TO: MAAN N

AMRL-TDR-62-98 (II)

AD 429681

TECHNIQUES OF PHYSIOLOGICAL MONITORING

VOLUME II. COMPONENTS

TECHNICAL DOCUMENTARY REPORT No. AMRL-TDR-62-98 (II)

NOVEMBER 1963

BIOPHYSICS LABORATORY
6570th AEROSPACE MEDICAL RESEARCH LABORATORIES
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

Contract Monitor: Mr. Miles A. McLennan
Project No. 7222, Task No. 722203

(Prepared under Contract No. AF 33(657)-9252 by
RCA Service Company, Cherry Hill, Camden 8, New Jersey)

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Qualified requesters may obtain copies from the Defense Documentation Center (DDC), Cameron Station, Alexandria, Virginia. Orders will be expedited if placed through the librarian or other person designated to request documents from DDC formerly ASTIA).

Do not return this copy. Retain or destroy.

Stock quantities available at Office of Technical Services, Department of Commerce, Washington 25, D. C. Price per copy is $4.00.

Change of Address

Organizations receiving reports via the 6570th Aerospace Medical Research Laboratories automatic mailing lists should submit the addressograph plate stamp on the report envelope or refer to the code number when corresponding about change of address.

1000 - December 1963 - 162-21-347

FOREWORD

This study was initiated by the Biomedical Laboratory of the 6570th Aerospace Medical Research Laboratories, Aerospace Medical Division, Wright-Patterson Air Force Base, Ohio. The research was conducted by the RCA Service Company, Camden, New Jersey, under Contract AF 33(657)-9252. Mr. Walter L. Becker was principal investigator for RCA Service Company. Mr. Miles A. McLennan of the Medical Electronics Section, Biophysics Branch, was the contract monitor for the Biomedical Laboratory. The work was performed in support of Project No. 7222, "Biophysics of Flight," Task No. 722203, "Specialized Instrumentation."

This is the second volume of a three-volume handbook. Volume I, prepared under Contract AF 33(616)-7750 and published in September 1962, gave a description of the basic physiological systems and their measurable parameters. Volume III, which is in preparation under Contract AF 33(657)-9252, will survey the various monitoring system configurations used in physiological monitoring.

This second volume was written by Mr. Richard Alnutt and Mr. Philip T. Weinberg, and edited by Mr. Robert E. Barbiere, all of RCA Service Company. Mr. Carl Berkley, Scientific Director of the Foundation for Medical Technology, participated as a technical consultant. Valuable comment also was received from Dr. Kurt S. Lion of Massachusetts Institute of Technology.

The authors thank the many manufacturers who supplied information during the preparation of this volume. While several manufacturers or their equipments have been cited specifically within the text, it was solely for the purpose of illustration, and such citation does not constitute endorsement by the authors, by RCA Service Company, or by the Air Force.

ABSTRACT

This volume surveys the components used in physiological monitoring systems, primarily those suitable for aerospace applications. Discussion includes performance characteristics and capabilities, plus some background theory, on basic components such as electrodes and transducers, signal modifiers, and graphic recording and display devices. The use of magnetic tape recorders in instrumentation is described. Wire and radio transmission equipment is discussed, plus various schemes of modulation and multiplexing. The capabilities of digital and analog computers and other data processing equipment are described, and the analysis of physiological data with such equipment is briefly discussed.

PUBLICATION REVIEW

This technical documentary report is approved.

Evan R. Goltra

EVAN R. GOLTRA, JR.
Lt Colonel, USAF, MC
Chief, Biophysics Laboratory

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

SECTION II - SIGNAL MODIFIERS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

SECTION IV – SPECIAL SYSTEM DEVICES

TABLE OF CONTENTS (Continued)

Page

SECTION V - DATA TRANSMISSION EQUIPMENT

SECTION VI - DATA PROCESSING EQUIPMENT

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

LIST OF ILLUSTRATIONS (Continued)

LIST OF ILLUSTRATIONS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

Section I

SENSING DEVICES

GENERAL

The electrodes and transducers discussed in this section are the most important links in the chain of components required for a physiological monitoring system. No matter how excellent are the other components in the system, the reliability of the data displayed or recorded is most critically dependent upon the quality of the signal originated by the sensing device. If the initial signal is faulty, it may be extremely difficult or impossible to remedy conditions by succeeding signal processing or modification.

Most physiological responses involve minute phenomena, small or delicate movements, low voltages, etc. The amount of power or energy involved is very small. The problem is to measure or display these phenomena without the act of measurement affecting and distorting the phenomena. Variations in body potentials and body impedances especially are minute quantities, necessitating extremely sensitive electrode systems and good contact between the electrodes and the subject. Physical displacement excursions are likewise difficult to detect without the sensing device physically interfering with the response to be measured. These restrictions place great demands on the design of the sensing device. At the least, a transducer must be extremely sensitive and small in mass.

The weak signals involved in physiological responses introduce another problem: that of interference. Displacement transducers may respond to more than one dynamic body output; electrodes may pick up more than one body potential; and movement of the subject during measuremunt may produce spurious outputs or completely nullify the established sensor-body interface. Extreme care therefore must be exercised in the attachment of sensing devices and in the interpretation of the outputs they obtain.

In the following pages, the factors of design, packaging, and attachment of electrodes and transducers are discussed. The basic principles by which the various transducers accomplish an electrical conversion are explained. The various configurations of electrodes and transducers are described; and their static, dynamic, and electrical characteristics are listed. Finally, the specific applications of the electrodes and transducers are discussed briefly, with emphasis upon their use in the environments encountered in the field of aerospace medicine.

Table I is a brief summary of the physiological parameters of interest, showing the type of sensing devices, employed in their measurement. These devices are all treated in the following text. Several references (ref. 21, 24, 34) are valuable as a guide to the full range of transducer capabilities; these books describe many devices which, while not at present in standard usage, have potential as biosensors.

TABLE I. BIOSENSOR SUMMARY

Measured Parameter	Measurement	Device	Sensor Used
Heart potentials.	Electrocardiogram	Electrocardiograph	Electrodes
Brain potentials.	Electroencephalogram	Electroencephalograph	Electrodes
Muscle potentials.	Electromyogram	Electromyograph	Electrodes
Skin resistance.	Galvanic skin response	GSR Recorder	Electrodes
Heart sounds.	Phonocardiogram	Phonocardiograph	Microphone
Blood flow.	Plethysmogram	1. Plethysmograph	1. Strain gage Differential transformer
		2. Impedance plethysmograph	2. Electrodes
Blood pressure.	Blood pressure	1. Sphygmomanometer	1. Strain gage Microphone
		2. Manometer	2. Strain gage Differential transformer
Pulse rate.	Pulse rate	Cardiotachograph	Microphone Strain gage Electrodes
Respiration rate.	Pneumotachogram	1. Pneumotachograph	1. Strain gage Thermocouple Thermistor Ionization tube Microswitch
		2. Impedance pneumograph	2. Electrodes
Volume of respired air.	Respiration depth	Spirometer	Multiplier photocell
Percent oxygen in blood.	Oxygen content	Oximeter	Photocell Polarographic electrode
Percent oxygen in respired air.	Percent oxygen	--	Electrochemical transducer
Temperature.	Temperature	Thermometer	Thermocouples Thermistor Resistance thermometer

ELECTRODES

Electrodes are simply conductors that serve as electrical paths for currents and potentials between parts of an electrical circuit. As used in physiological monitoring, electrodes have two distinct applications: one of these is to connect electrical potentials generated in the body to external measuring instruments; the other is to impress externally generated currents across portions of the body for the measurement of changes in body impedances.

Body potentials and impedances can be measured both at and beneath the surface of the body. Most of this discussion is concerned with surface electrodes, because subcutaneous and surgically implanted electrodes have limited application in aerospace environments. The following paragraphs discuss the various types and configurations of surface electrodes, the electrode materials and electrical characteristics, and body attachment factors. (Additional material of value on electrodes is contained in ref. 49.)

I. General Electrode Characteristics

A. Sources of Body Potentials

Essentially, there are two types of electrical phenomena measured at the surface of the body: deep seated and surface. Signals of both origins are minute, with potentials ranging from a few millivolts for cardiac potentials to about 20 microvolts for brain waves. Significant impedance changes that can be measured at the surface, with the exception of baseline skin response measurements, generally are below one percent.

Sources of deep body potentials include the heart muscle potentials, the measurement of which is called electrocardiography (ECG), and potentials from the brain or central nervous system, the measurement of which is called electroencephalography (EEG). Action potentials from skeletal muscle also may emanate from deep tissue: their measurement is called electromyography (EMG). As measured in most aerospace applications, they are near-surface phenomena.

The other surface electrical phenomena measured are various forms of body impedance. These include the galvanic skin response (GSR), a near-surface phenomenon in which changes in skin layer impedance are measured as a function of behavioral condition. Also included are changes in body impedance as a function of respiration (the impedance pneumograph) and regional blood flow (the impedance plethysmograph).

B. Electrode Materials

Several factors must be considered when selecting electrodes. The electrode must be a good conductor to permit good current transfer between the subject and

instrumentation, and, to minimize electrochemical artifacts, it should be inert to chemical reactions. Further, electrodes should be able to withstand the environmental conditions of the measurement and still give unhindered or relatively stable operation.

Depending upon the application, there are various configurations of metal electrodes: plates or discs, needles, and meshes. Other configurations, less often used, are glass tubes filled with electrolyte, and sponge materials impregnated with electrolyte. The glass electrodes are basically biological research tools, where the electrical activity being studied is molecular or cellular in nature. The sponges are frequently employed in the measurement of galvanic skin response.

Metal electrodes are used either dry, applied directly to the body, or wet, where the electrical contact with the body is reinforced by the application of an electrolytic paste or jelly. In such usage, the combination of paste and metal contact is considered the electrode.

The significant characteristics of electrode materials are discussed below.

1. Conductivity

A good conductor is a material that offers little resistance to the flow of an electric current. The electric current may be the movement of electrons, as it is in metals, or it may be the migration of ionized atoms or molecules, which takes place in gases and liquids. In either case, the conductivity of the material is determined by the quantity of electricity transferred across a given area of the material per unit of potential gradient per unit of time.

The voltage potential (and consequently the signal current) obtained from biological tissue is relatively low (usually about a few millivolts); therefore, the electrode must offer a minimum of resistance to current flow in order to minimize signal attenuation between the source and the monitoring instrumentation.

2. Inertness

Inertness, or the relative inactivity of a material in chemical reaction, is a desirable characteristic for electrodes. If not relatively inert, the formation of ions by the electrode material, and the combining of these ions with enzymes, produces poisons which can irritate the skin of the subject.

Ions are formed when electrodes dissolve in acid body fluids, such as sweat, and the electrode molecules replace (give up electrons to) the hydrogen ions present in the fluids.* The degree to which electrode materials are subject to chem-

*The dissolution of electrode materials in fluid is itself objectionable, since it can in time reduce effective skin-electrode contact through a decrease in the amount of electrode material.

ical reaction varies. Table II lists the major metals used as electrodes in their order of chemical activity. The more active metals, such as zinc and nickel, readily replace hydrogen; those less active than hydrogen, such as silver and gold, do not. The hydrogen ions themselves partake only minimally in objectionable chemical reaction.

TABLE II. ELECTROCHEMICAL SERIES

Magnesium	Lead
Aluminum	Copper
Zinc	Mercury
Iron	Silver
Nickel	Platinum
Tin	Gold

3. Polarization

When a current is passed through a metallic electrode immersed in a solution or inserted into biological tissue, a phenomenon known as polarization occurs. Regardless of how inert a metallic conductor may be, it still participates in a certain amount of chemical and physical reactions. The fact that all metals are soluble (however slightly) and form ions results in a d-c boundary or polarization potential between the electrode and the fluid establishing contact with the electrode. This polarization potential introduces artifacts into the physiological measurements and thus tends to mask the information being monitored.

Although there is no truly nonpolarizable electrode, the general practice of using electrode materials that are as insoluble and nonreactive as possible reduces the tendency to form ions and thus minimizes polarization. As a result, the noble metals (silver, platinum, and gold) are often used for electrodes. Chemically active metals like zinc and nickel tend to introduce a large polarization potential and, consequently, excessive errors in physiological measurements.

When measurements are to be made for extended periods, the metal electrodes used should be composed of special alloys that retard electrolysis and minimize the effects of polarization. A silver electrode with a thin coating of silver chloride deposited upon its surface has been used successfully, as well as combinations of platinum and platinum chloride and zinc and zinc sulphate. Note, however, that junctions of dissimilar metals should be avoided where skin contact is made. Even a soldered joint may be troublesome.

4. Impedance

A typical resistive component of impedance, as measured between the leads of a pair of one-square-inch metal electrodes attached to the skin, is about 5000

to 100,000 ohms. Careful skin preparation and the use of electrolytic paste or jelly can reduce this figure to below 1500 ohms, at least for measurements of no more than a few hours duration.

II. Electrodes for the Measurement of Deep Body Responses

The measurement of deep body tissue responses, such as the ECG and EEG, is possible with most of the known configurations of surface electrodes. Criteria for selecting electrodes for a particular application include the fidelity of signal desired, the length of time of the measurement, and the degree of subject movement that can be tolerated. All of these factors relate to how large an electrode must be employed, and how much care should be taken in preparing the skin and attaching the electrode.

A. Metal Plates

The most rudimentary form of surface electrode is a thin metal plate or disc, to which lead wires are attached with solder. This electrode is durable and easily attached with tape, straps, or adhesives. With a large area of metal in contact with the skin (optimum is under 2 square inches), the impedance of the attachment is low enough for adequate electrical contact, even when the skin-metal contact is dry. However, the large area of the electrode (and its high mass) makes it subject to movement artifacts, such as the spikes of myographic potential that accompany movement. Where subject movement is required or permitted during measurement, metal plates are best kept small and employed in a floating configuration (floating electrode), in which they are restrained in a suitable adhesive holder, but prevented from directly contacting the surface of the subject by a cushion of electrode paste or jelly.

Figure 1 shows typical metal plate electrodes. Electrodes of this type have been fabricated of silver, platinum, and even stainless steel.

B. Meshes

Small meshes of wire, either circular or square, also are used as metal electrodes. The mesh configuration provides for good contact between the electrode and electrode jelly. Metal meshes vary in area from one-fourth square inch to an inch or more. The smaller sizes usually yield better results with regard to movement artifacts.

Two electrodes of this type that are available commercially are shown in figure 2. One electrode has a small, square wire grid, fastened to an adhesive tape backing, with a snap-on electrical connector on the back. The grid is coated with a suitable electrode jelly before being taped to the skin. This electrode will pick up movement artifacts, but, where this is not a consideration, it offers the advantage of quick attachment and easy electrical connection, and, being relatively inexpensive, it is disposable after use.

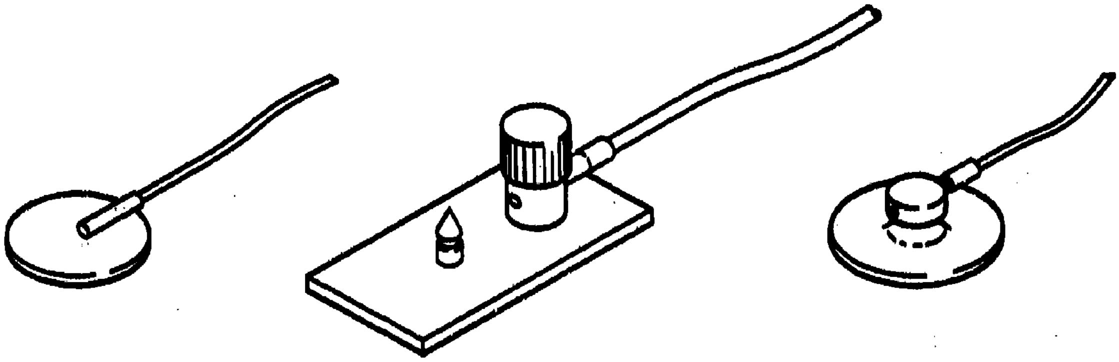
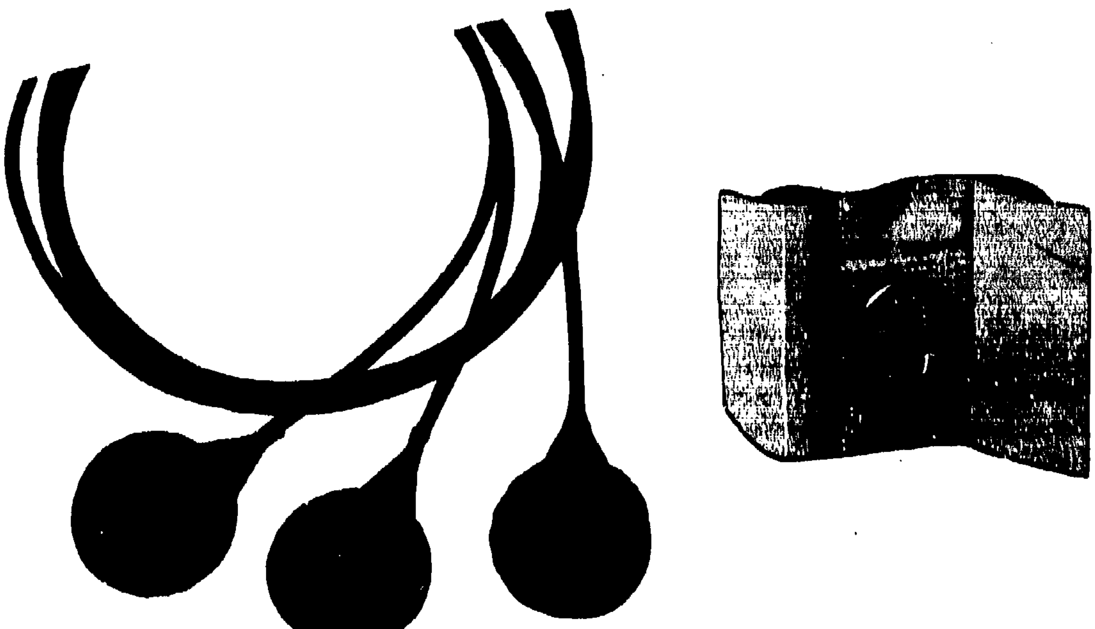


Figure 1. Typical Plate Electrodes



Cordis Corp., Miami, Fla.

Telemedics Inc., Southampton, Pa.

Figure 2. Typical Mesh Electrodes

Circular meshes or grids also are available in flexible housings, such as the molded assembly shown in figure 2.

C. Metal Cups

A third form of metal surface electrode is the cup. These small, hollow electrodes usually are fabricated of silver. One type used in hospitals or clinics for ECG and fetal ECG measurements is the Welsh vacuum cup. It is placed against the skin, and a rubber bulb is used to expel the air from within the cup. This draws the skin up into the cup, and holds the cup firmly against the skin. The electrical connection is good, but the vacuum is short-lived, making this type of attachment unsuitable for long-term measurements.

Other cup electrodes are used in aerospace environments with success, principally for EEG measurements. This type of electrode is attached to the skin with adhesive (cement, collodion, etc), and the hollow portion is filled with electrode jelly (see figure 3). A good liquid contact is achieved, and it can be maintained for extended measurements of up to 24 hours, because the electrolyte is under a near-hermetic seal. With careful preparation of the skin, electrode lead impedances under 2000 ohms can be attained.*

D. Needles

Small steel needles, placed just beneath the skin, make excellent electrodes for ECG or EEG. They have much higher impedances than electrodes attached on the surface, but they are much less affected by movement artifacts. Needle electrodes are little used, however, because of subject discomfort and irritation to the tissue during prolonged measurements.

E. Metallic Cloth

Metallic cloth electrodes, either in thin patches or thick pads, are composed of cloth woven of fibers coated with a thin film of metal; nylon-silver is a common type. These electrodes are useful for long-term measurements. They are flexible, providing large-area contact and easy conformance to body contours, and are well suited for attachment through biosensor harnesses. Cloth electrodes generally are applied without electrode paste, resulting in a high-impedance contact. If used for ECG measurements, then, such electrodes would require the use of preamplifiers with high input impedances.

---

*From a series of in-flight electroencephalographic measurements conducted at Wright-Patterson Air Force Base under the direction of Dr. C. W. Sem-Jacobsen.

Figure 3. Use of Collodion With Small Cup Electrode

III. Electrodes for the Measurement of Surface Responses

Measurement of surface electrical phenomena, with the exceptions described below, are possible with various configurations of the electrodes used to measure deep body tissue responses; however, differences in impedance, bandpass frequency, etc, must be compensated for in external circuits. In one application both ECG and impedance pneumograph measurements can be made simultaneously, using the same electrodes. Electrode configurations particular to the measurement of surface phenomena are described below.

A. Metallic Cloth Pads

High-impedance electrodes fabricated of metallic cloth (nylon-silver) are suitable for measuring galvanic skin response. The cloth is folded over, filled with padding, and stitched into a soft pad, about 1 x 2 inches in size, as shown in figure 4.

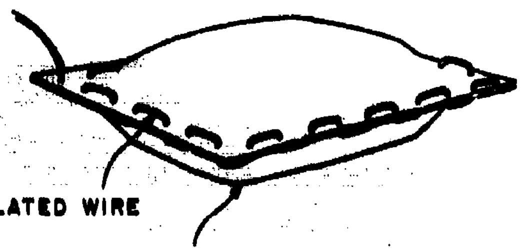


Figure 4. A GSR Electrode of Metallic Cloth

The lead wires may be used for the stitching. The electrode is attached dry to the body (the palm of the hand or sole of the foot) with adhesive tape or straps.

B. Soft Metal Plates

Plates of soft metal, such as lead, sometimes are used as electrodes for the GSR measurement. They are placed against the body dry and held in place with tape or straps. Soft padding is used between the tape and the electrode to ensure compliance with the body by distributing the pressure from the tape more effectively.

C. Electrolyte-Filled Sponge

GSR measurements also are made using soft sponges, the cells of which have been impregnated with an electrolytic solution. This type of electrode makes excellent electrical contact with the skin, but is not suitable for measurements in an aerospace environment. The application is messy, and it is difficult to cover the electrode adequately so that the electrolyte does not evaporate. When the electrolyte does evaporate, the electrical signal degrades sharply.

D. Ring Electrodes

Electrodes in a ring configuration are attached to the finger for blood flow measurements called impedance plethysmography. The electrodes are placed over the fingers dry, and are held in place by skin friction. The better conductors, such as gold-plated electrodes or electrodes of silver or monel, should be employed but success has been reported in this measurement using almost any available metal (ref. 33).

IV. Attachment Factors

Several factors, including physical placement, skin preparation, and electrode adhesion, must be considered when bioelectric potentials are measured on the surface of the skin. The location of the electrodes on the subject's body depends on such considerations as proximity to the signal source, remoteness to contaminating noise, and the requirements of not producing irritation or limiting normal body movements. The problem of maintaining a constant skin resistance for extended periods of time must be coped with when preparing the skin for electrode attachment, and, when physically attaching the electrode to the body, care must be taken not to change the subject's blood circulation or produce subject discomfort. The electrode must be attached in such a manner as to maintain a constant pressure on the skin surface.

A. Electrode Placement

The problem of electrode placement naturally revolves around the type of information (electrocardiogram, electroencephalogram, galvanic skin response, etc) sought. Supplementing this controlling factor are three prime considerations: (1) the interpretability of the information obtainable from the various possible body sites, (2) the signal-to-interference ratio obtainable from these sites, and (3) the physical restrictions imposed by the placement of the electrodes.

Electrodes placed on the extremities produce a fairly large signal for some applications; nowever, the interference level resulting from most of the standard body movements often presents an acute problem. The amount of muscle mass underlying different body surface areas is another important factor to consider, since muscle potential produces interference and artifacts.

B. Skin Preparation

Since a constant skin resistance should be maintained when obtaining physiological measurements, the variables introduced by the characteristics of the subject's skin must be minimized by properly preparing the skin before attaching the electrodes. Most important in skin preparation is that the skin be cleansed thoroughly. Since ordinary washing removes only the more water-soluble materials, a fat solvent (ether, alcohol, acetone, etc) may be necessary to remove the salt solution and the oily layer produced by the sweat glands and the sebaceous glands, respectively. Mild detergents also are recommended because of their ability to penetrate the skin and remove the oily matter; however, these compounds leave the skin wet. If the area is dense with hair, shaving may be necessary prior to cleansing.

A slight abrasion of the skin also may be useful, in short-term work, to remove the cellular debris and dried salt present on the surface of the skin. This abrasion should not be deep, but merely to the point of causing a slight reddening because of the increased blood flow to that surface. If abrasion of the skin is not desirable (as in

GSR measurements), a conducting jelly may be rubbed into the skin to permit direct electrical contact with the low resistance component of the inner layer of skin.

C. Adhesives

The materials used for attaching electrodes to the skin are of prime importance. Even an ideal electrode placed on properly prepared skin obtains a signal contaminated with artifacts if the attachment is unsatisfactory. A properly attached electrode must (1) maintain a constant pressure on the skin surface, (2) not inhibit movement or limit blood flow, and (3) not make the subject unpleasantly aware of the electrodes. These three requirements should be met during the entire time that measurements are being taken.

Ordinary adhesive tape is a simple attachment material that can be used satisfactorily with most electrodes having flat surfaces. The two primary advantages of using tape rather than some other form of adhesive material are its simplicity and economy. One of the main disadvantages is that tape may fail to maintain contact pressure if the skin is moist with saline solutions. Also, large pieces of tape often are required, and thus may inhibit motion or overlap into other moving areas. Adhesive tape adheres best if the area to be taped is coated with tincture of benzoin or with rubber cement.

Many commercial cements and pastes are available for attaching electrodes to the skin. Adhesives used for this purpose should be quick drying and must adhere reasonably well to the skin. Adhesive cements and pastes prevent salt bridges (lowered resistance pathways) between electrodes. In addition, it is possible to achieve sealed connections that result in the enclosed area remaining moist and the skin resistance remaining constant.

Certain precautions are necessary when an adhesive cement or paste for electrode attachment is selected. Some plastic-type adhesive cements, for example, become rigid and brittle when dry and crack easily. Other adhesives may contain volatile solvents, which may be irritating to the point of discomfort or have noxious odors.

D. Mechanical Devices

Special devices are sometimes necessary for electrode attachment to comply with anatomical contours and limitations. For short term applications, plastic or rubber belts often are used, since they are applied easily and are capable of maintaining a constant pressure. These and other devices, including slings, gloves, and sleeves, may be used alone or with electrode adhesives; however, mechanical devices should be used with discretion in long-term measurements, since they tend to create subject discomfort.

E. Floating Attachments

An electrode configuration that has been used with much success in aerospace applications is the so-called floating electrode. Developed over the last few years, primarily for airborne ECG recording, this electrode consists of a sandwich of layers of cork, within which a metal disc or mesh is floated in a restrained volume of electrode paste or jelly. This type of electrode, which produces measurements remarkably free of movement artifacts, can be made small in size and light in weight, and it is somewhat flexible, so that subject discomfort is minimized.

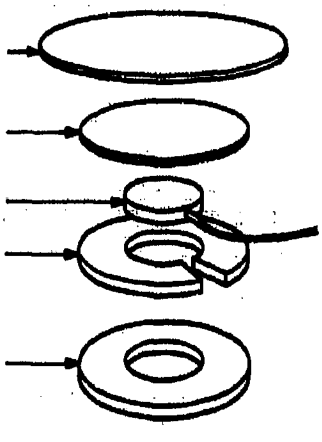


Figure 5. Configuration of a Floating Electrode

Figure 5 shows a floating electrode developed at the Naval Missile Center, Point Mugu, California (ref. 25). This electrode uses three discs of cork tape (lower disc, slotted retaining ring, and top disc) to restrain the metal disc, which is fabricated of equal parts of silver and silver chloride, compressed into a pellet at 20,000 pounds per square inch pressure. This type of electrode can be attached in about 5 minutes as follows:

The skin is first cleaned and then abraded with the electrode jelly, after which it is treated with tincture of benzoin (except for the area under the center hole). The lower disc, with slotted ring attached, then is pressed in place. The skin within the center hole is given a thin coating of electrode jelly, and the hole is filled with a jelly-impregnated sponge cylinder. The underside of the metal pellet is coated with jelly and placed in the center hole, with the lead wire seated in the slot. The top disc is pressed into place, and finally the cover patch is applied over the whole electrode.

This electrode provides an impedance of between 500 and 2000 ohms, well below the norm for ECG electrodes. It was tested up to 24 hours, and there was

neither significant degradation in performance nor significant irritation to the skin of the subject.

Electrodes of this type are fabricated readily, or they may be obtained commercially. Conformance to the performance figures cited above will depend largely upon the size of the components, the care with which the skin is prepared, and the type of electrode jelly employed. (The floating configuration affords a near-hermetic seal, minimizing evaporation of electrolyte, but this will vary with the type of volatile materials employed in making the jelly or paste.)

F. Surgically Implanted Electrodes

To obtain physiological potentials from deeply rooted structures (brain tissue, muscular tissue, etc) surgically implanted electrodes are used instead of the ordinary surface electrode. Implanted electrodes afford better localization and access to the deeper-lying structures, they encounter less resistance, and the signal they obtain is not contaminated by the galvanic skin response or other d-c or low-frequency waves from the adjacent body areas. Surgically implanted electrodes also are referred to as subcutaneous, penetrating, needle, subdural, intramuscular, and subcortical electrodes.

1. Requirements

Surgically implanted electrodes must meet the following requirements if they are to provide optimum results: (1) they must be rigid enough to provide mechanical penetration, (2) they must be shaped in a manner that facilitates insertion with a minimum amount of damage to body tissue, and (3) they should be well insulated to limit the area being monitored to that which is in direct contact with the uninsulated tip.

2. Basic Configurations

Enlarged, long-axis cross-sectional views of some of the basic implanted electrode configurations are shown in figure 6. A simple surgical needle, usually made of stainless steel, is shown at the top. This type of electrode normally is short and connected to a conducting wire. The same basic needle with an insulating material covering all but the tip, confines the area of measurement to that which makes contact with the needle tip, and better localization is provided. A standard hypodermic syringe needle with a conducting wire cemented in it for stability also is shown. This type of configuration permits the needle shaft to be used as a shielding device, provided the cement which is used is a good insulator. However, sharp penetrating edges must be maintained on the syringe.

3. Restrictions and Required Precautions

Implanted electrodes present such problems as subject discomfort, trauma to the tissue, sterility requirements, and the possibility of penetrating blood vessels and

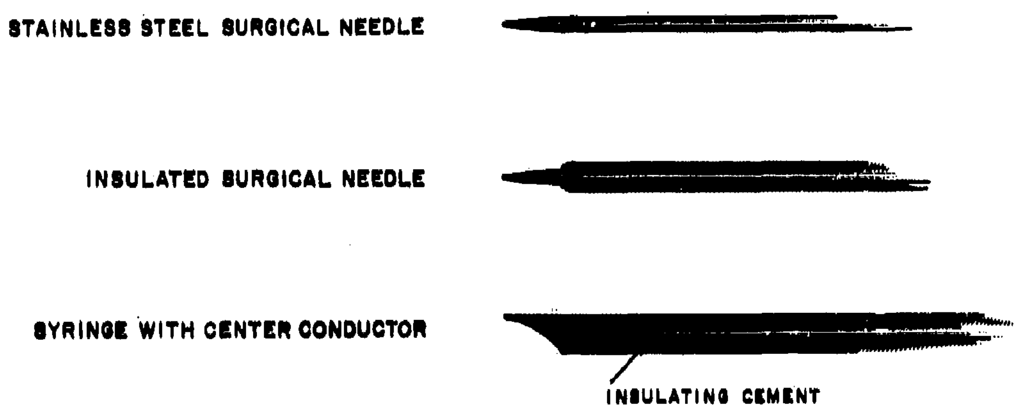


Figure 6. Basic Implanted Electrode Configurations

nerve fibers. As a result, implanted electrodes are used mostly in test laboratories. Much care and skill is needed to cope successfully with the problems associated with electrode implantation.

The surfaces of implanted electrodes should be highly polished before insertion to prevent unnecessary tissue damage. Any insulation material on the electrode should be nontoxic and "nonwettable." The use of nonwettable insulating materials prevents the formation of salt bridges and minimizes galvanic generation of extraneous d-c voltages.

DISPLACEMENT TRANSDUCERS

1. General

Electrical conversion in most mechanical displacement transducers depends upon the movement or displacement of a sensitive member. Auxiliary levers or couplings transfer the movement from the biological specimen to the sensitive member. Other phenomena, such as force, pressure, and acceleration, are detectable by displacement transducers, but must first be converted to a displacement by suitable mechanical restraining and coupling devices.

The mechanical displacement transducer is used to detect physiological responses in the cardiovascular and respiratory systems, as well as to provide information of environmental surroundings. In cardiovascular applications, the transducer monitors blood pressure, blood pulse wave signals, phonocardiograms, and ballisto-vibrocardiograms. In respiratory applications, it monitors respiration rate and respiration depth measurements.

A. Blood Pressure Measurements

Blood pressure may be measured either directly or indirectly. In the direct methods, an incision into the subject's body (usually a limb) is made, and a transducer or a transducer transmission device is inserted within the blood vessel. A catheter usually is inserted into the brachial artery, and the transmitted pressure profile is impressed upon a diaphragm composed of a sterile, noncorrosive material. The diaphragm is linked mechanically to a displacement transducer.

An even more direct method of measuring the blood pressure is by inserting the transducer directly into the blood vessel. This may be done by constructing a probe using a standard hypodermic needle as the shell for the transducer.* The pressure head within the artery acts on the transducer through a diaphragm inside the hypodermic. In this manner the losses involved in transmitting the blood pressure profile through the catheter are eliminated, providing better frequency response.

Measuring blood pressure indirectly with a displacement transducer is possible by the sphygmomanometer or "cuff and microphone" method. Here the transducer is simply a microphone which detects the cessation or continuation of blood flow in an occluded vessel.

B. Blood Pulse Wave Signals

Displacement transducers are used for monitoring pulse wave profiles. The pulse measurements are made by detecting the minute deflections of body tissues caused by the pulsation of the blood within the blood vessels along a limb or at a fingertip. The profile of the pulse wave may be correlated with other data, such as impedance plethysmograph information, to obtain blood pressure measurements and pulse rates.

C. Phonocardiographic Detection

In this application, the displacement transducer, used as a microphone, is fastened to the subject's chest in the heart region, and the output of the microphone is recorded or instantly displayed upon an oscilloscope. This type of analysis of heart sounds may contain as much characteristic information as an electrocardiogram. A typical device of this type provides a frequency response of 20 to 1000 cycles per second.

D. Ballisto-Vibrocardiogram Measurements

Displacement transducers are used as very low frequency accelerometers to detect the movement of blood in the body. The subject is placed on a ballistic table,

---

*Biometrics Instrument Corporation, Dallas, Texas

isolated from environmental shocks and vibrations by a carefully designed spring suspension system. The movement of body mass from the flow of blood through the circulatory system produces detectable accelerations, which are measurable by displacement transducers.

E. Respiration Measurements

The displacement transducer often is used to measure the expansion of the subject's chest and provide a respiration rate (pneumotachograph) and depth measurement. The expansion of the chest is absorbed by an elastic band around the chest. Inside the band is a displacement or strain gage type of transducer that is stressed in an amount proportional to the entire expansion of the chest.

II. Piezoelectric Devices

Piezoelectric devices are used as transducers because an electric charge is produced across the face of a type of solid crystalline dielectric when the crystal lattices of the dielectric are distorted in a certain manner. The elasticity and thickness of the crystal determine the amount of charge produced by a force on the crystal by the following relationship:

$$Q = dF$$

where

Q = the amount of charge produced by force F

d = constant of proportionality (coulombs/newton or $\frac{\text{coulombs/m}^2}{\text{newtons/m}^2}$).

The piezoelectric crystal may be thought of as a capacitor with the charge appearing across the two parallel plates. The value of the capacitor is directly proportional to the area of the planar faces and inversely proportional to the thickness of the slab of crystal used. Thus the voltage appearing across the crystal faces is the charge Q divided by the capacitance C or

$$V = Q/C$$

$$= \frac{dF}{\frac{\epsilon A}{t}}$$

where

ϵ = the dielectric constant

A = the area of the crystal

t = the thickness of the crystal slab

Calling d/ϵ the voltage sensitivity for a given crystal cut in a particular manner g, and considering the force per area of the crystal F/a or pressure P, an expression relating the voltage output of the crystal in terms of the pressure applied is obtained:

$$V = gtP$$

Values for g range from -0.29 to +0.354 $\dfrac{\text{volts/m}}{\text{newtons/m}^2}$

The equivalent circuit for the piezoelectric crystal shown in figure 7, indicates the limitations to its use. The charge produced by the mechanical stress on the crystal may be considered to accumulate on capacitor C. Leakage resistor R tends to shunt out this voltage with a time constant RC. However, external load impedances are usually lower in magnitude than the internal resistance, and these external loads usually determine the amplitude and frequency response of the crystal output. An upper frequency limit near the limit of the audio band may be achieved. Because of the blocking effect of the crystal capacitance, a static d-c response is not transmitted. The piezoelectric crystal is not linear throughout its range, but does exhibit linear responses within confines of pressure limits.

The device is continuous in response and thus provides infinite resolution. The sensitivity of the device varies with the composition and the cut of the crystal, as indicated by the wide spread of values of g. The output level sensitivity for a typical unit* is 93 db below 1 volt per dyne per square centimeter.

Some recent developments in artificially produced piezoelectric crystals have produced extremely high voltage output devices in the order of kilovolts.

III. Variable Resistance Devices

Devices which use mechanical movement to change an electrical resistance are used to sense displacements of body structures and to measure forces produced by the physiological processes in the body.

A. Strain Gages

The strain gage is most useful where very small displacements are concerned and may be calibrated in terms of linear displacement, or the force, pressure, or acceleration involved to produce the displacement. Strain gages are of three types: wire, semiconductor, and electrolytic.

1. Wire Strain Gage

There are two types of wire strain gages: unbonded and bonded. With both types, the displacement is measured by observing the degree of change in the resistance of a segment of very thin wire when it is subjected to displacement force. The wire used usually is composed of a nickel and copper alloy, iron, chromium, platinum, or constantan.

*Gulton Industries, Inc., Metuchen, N. J.

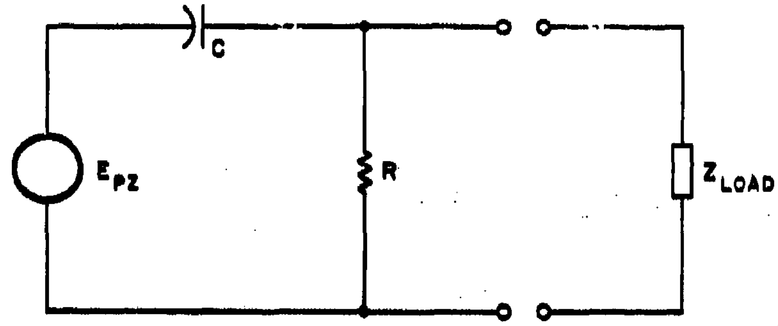


Figure 7. Equivalent Circuit for a Piezoelectric Crystal

An unbonded wire strain gage is assembled as in figure 8. In this configuration, the transducer is in the form of a Wheatstone bridge. A displacement force on the gage strains two of the arms and relieves the other two, giving a strain gage output with the relation

$$\frac{\Delta R}{R} = S \frac{\Delta L}{L}$$

where

ΔR = the change in the resistance from the unstressed value R,
ΔL = the change in the length of the gage from its unstressed value L
S = the constant of proportionality, is the sensitivity factor.

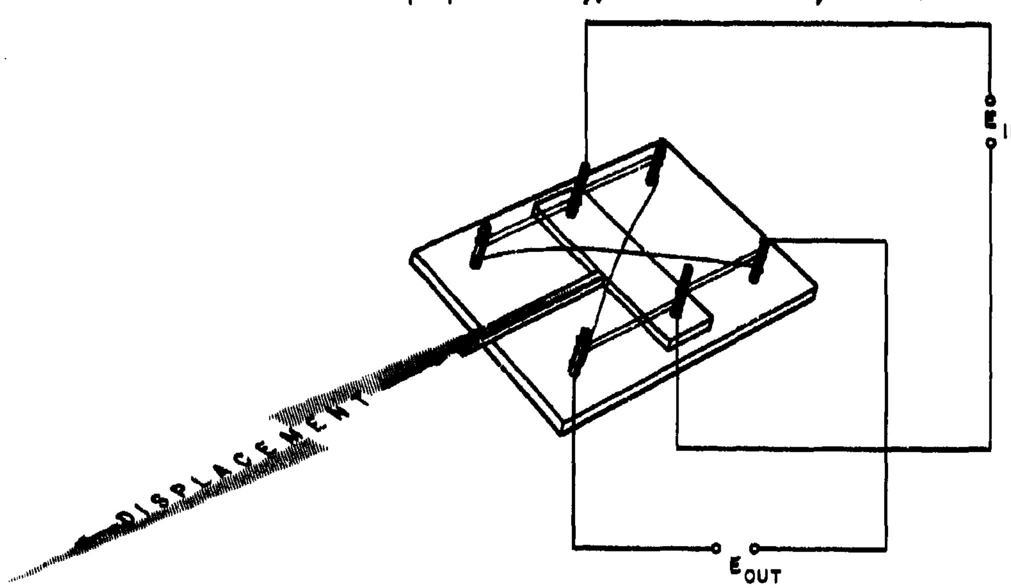


Figure 8. Unbonded Wire Strain Gage

A high value of S indicates a greater resistance variation for a given strain, with subsequent reduced demands for amplification and noise filtering, which is a desirable situation. Some typical values for the sensitivity factor follow:

Nickel-Chromium	2.1 to 2.3
Nickel	-12.1
Constantan (Nickel-Copper)	2.0 to 2.1

The bonded wire strain gage consists of a wire bonded to a backing of paper or epoxy with a cement or rosin. The wire pattern is some type of a grid; or a foil material is used instead of a wire as the stressed element. The materials used for the stressed element are the same as in the unbonded gage. The backing paper is cemented in position over the stressed or moving member. The expression for the response of the bonded strain gage is

$$\frac{\Delta R}{R} = G\epsilon$$

where ΔR and R are as defined above, ϵ is the applied strain, and G is the proportionality constant, called the gage factor. The gage factor has essentially the same significance as the sensitivity factor for the unbonded gage. Values for G of some typical units* follow:

Wire Grid - Constantan	2.1
Wire Grid - Iso-Elastic	3.3
Foil - Constantan	2.1
Foil - Nichrome	2.1

The size of the wire strain gages depends mainly on the size of the backing material to which they are bonded. The smaller the area of the strain gage, the more accurate is the measurement for point strain determination. For most physiological experiments, the size of the strain gage may be made compatible with the body parameter under measurement. The weight of the wire or foil is negligible compared to the backing material, and the whole strain gage weighs less than an ounce.

The frequency response for bonded wire strain gages may extend from dc to many thousand cycles. The usual ranges, however, are from dc to a few hundred cycles (ref. 24).

Wire strain gages may be obtained at any nominal resistance value. Values commercially obtainable "off-the-shelf" range from 60 to 100 ohms to as high as 4000 to 5000 ohms. Thus a low- or medium-impedance system may be used.

The variation in the nominal resistance with force or displacement applied to the wire strain gage is in the order of 1 to 5 percent. In most applications, this resistance variation changes the output in the millivolt range, depending, of

*Baldwin-Lima-Hamilton Corp., Waltham, Mass.

course, on the strain gage excitation voltage. Self-heating of the devices by the excitation voltage applied must be avoided. If possible, to allow complete stabilization of the transducer it is best to wait a few seconds between the time that excitation is supplied initially to the strain gage and the time a determination of its nominal value is made.

Environmental temperatures limit the use of the strain gage for displacement and force measurements. Since a resistance change in the gage is the parameter being monitored, changes in the gage resistance due to external factors, such as ambient temperatures, cannot be tolerated without compensation.

Temperature variations may be compensated for in two ways: another identical gage, unstressed, can be exposed to the same environment and used as a reference for the nominal resistance, or another material can be used in the gage construction whose temperature coefficient is such that it cancels the temperature coefficient causing the error. In addition, the stresses resulting from the difference in the coefficients of expansion between the strain gage material and the subject under test must be compensated for. In the environmental temperatures to which the human body is exposed, temperature compensation is not a significant problem. Also, because the wire strain gage is used most often for dynamic measurements, a shift in the baseline does not often affect the usability of the data.

2. Semiconductor Strain Gage

In the semiconductor strain gage, the stress is concentrated on a single filament of silicon material, doped to the desired impurity level.* When this filament is stressed, its internal resistance changes in a manner similar to that of the wire strain gage. Since the gage factor for semiconductor strain gages is in the order of 120, compared to 2 to 3 for wire gages, this type gage is much more sensitive. The gage factor, G, is defined in the same manner as for the bonded wire gage,

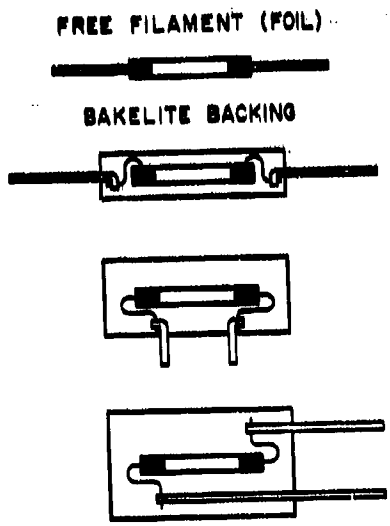
$$\frac{\Delta R}{R} = G\epsilon$$

*Doping of a semiconductor crystal is the term used to describe the intentional introduction of impurities into the basic silicon (or germanium) melt to produce the desired level of resistivity in the material.

The semiconductor strain gage has two main drawbacks: the device is somewhat nonlinear throughout its range, and its resistive values are sensitive to temperature. When used in a bridge configuration, this temperature sensitivity may be minimized.

Semiconductor strain gages, like the wire strain gages, are fabricated in both a bonded and unbonded configuration. Since they are rather fragile, the semiconductor gages usually are bonded for reliable operation.

The frequency response of the semiconductor strain gage is similar to that of the wire gage: dc to thousands of cycles. Their small mass and size make them useful for measuring point sources. Typical semiconductor strain gages* are illustrated in figure 9, with both bonded and unbonded free filament types shown.



Baldwin-Lima-Hamilton Corp., Waltham, Mass.

Figure 9. Typical Semiconductor Strain Gages (Bonded and Unbonded)

*Baldwin-Lima-Hamilton Corp., Waltham, Mass.

3. Electrolytic Strain Gage

A third type of strain gage is a device fabricated specially for physiological research. This device, the electrolytic or liquid strain gage, measures changes in volume as opposed to the linear changes measured by the two gages discussed previously.

The liquid strain gage usually is constructed of a length of thin flexible tubing, filled with a liquid capable of conducting an electric current. The most commonly used substances are mercury and electrolytes such as copper sulfate, zinc sulfate, and silver nitrate. The tubing is filled with the solution and stoppered at each end with conductive plugs made of silver or platinum for mercury, or the metal of the electrolyte, copper, zinc, or silver. The total resistance from end to end of the tubing, measured through the conductive plugs, is approximately

$$\frac{\rho l}{\pi r_i^2}$$

where

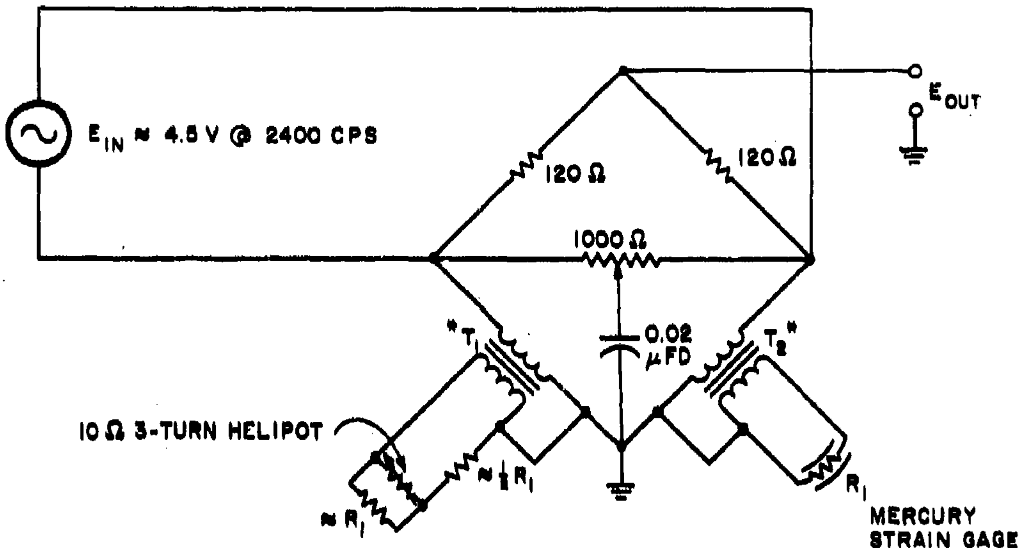
ρ = the resistivity of the liquid conductor
l = the length of the rubber tubing
r_i = the inside radius of the tubing

Since the stress forces imposed on the device are mainly in a direction changing the length of the tubing, the measured resistance varies according to the elongation imposed on the tubing by physiological stress forces.

If mercury is used as the conductive fluid, a resistance of less than 5 ohms is obtained for a length of tubing useful in plethysmography. For a low resistance gage such as this, a special impedance matching circuit (ref. 12) is necessary, such as is shown in figure 10.

An electrolyte liquid strain gage may have a resistance as high as 20,000 ohms or as low as a few hundred ohms, depending on the concentration of the electrolyte. With electrolyte and metal electrodes, electrolysis occurs with subsequent erosion of the terminal contacts. This dissolution of the electrode material can be minimized by the reversing potential of the a-c field across the electrodes when an a-c generator is used as in the illustrated impedance bridge. If the fluid strain gage is used in a d-c configuration, the dissolution of the contacts must be minimized either by periodically reversing the current through the gage or by using a current through the gage so small that the electrolysis produced is negligible compared to the desired life of the transducer (ref. 26).

The liquid strain gage does not have the extended frequency response of the wire and semiconductor gages because of its increased size and mass. The response



Legend:

*T_1, T_2 = CHICAGO STANDARD TRANSFORMER CO., MODEL A-8103

Figure 10. Impedance Matching Circuit for the Mercury Strain Gage

extends from dc to about 100 cycles per second. The gage may be made more rugged to withstand environmental abuse by enclosing the rubber tubing in a light metal spring or some similar protective outer shell that does not interfere with the sensing function of the device.

B. Potentiometric Devices

Potentiometric devices also are used as variable resistance transducers in physiological monitoring systems. In this device, mechanical motion is translated, through a linkage, to movement of a slider arm of a potentiometer, changing the potentiometer resistance. The potentiometer may or may not be linearly constructed; a nonlinear construction may be used to compensate for nonlinearities in the mechanical portion of the transducer system.

The linkage system may be a pure displacement type arrangement with the pressure of the slider arm of the potentiometer on the slide wire or resistive material inside the potentiometer the only force to be overcome. Other devices use a cable that is spring loaded to keep it wound up; mechanical movement unwinds the cable and subsequently moves the slider arm of the potentiometer. For pressure measurements, a Bourdon tube may be linked mechanically to the potentiometric device, with the two elements in a single housing.

The potentiometric transducer can produce a strong, varying signal in response to a small mechanical movement by using a large excitation voltage. The limitation to the amplitude of the excitation voltage is the amount of power the device can dissipate, which is determined by the resistance range of the potentiometer. Potentiometers are constructed with a total resistance in the order of a few ohms to many thousands or even millions of ohms. The potentiometer may be used in conjunction with a voltage or current amplifier or it may be used in a bridge circuit to drive telemetering or recording equipment directly.

The frequency response of the potentiometric transducer is limited by the mechanical linkage associated with the transducer and the mass of the wiper arm of the resistance element itself. Most of the transducers used in this configuration respond to less than 100 cycles per second. The resolution of most of the available devices is better than 0.005 inch. The linearity of the potentiometric transducers is better than 0.6 percent for many units (ref. 24).

C. Carbon Granule Devices

The carbon pile or microphone is the third main classification of variable resistance devices used to transform motion or force into an electrical parameter. In this device, small granules or discs of carbon or a carbon-binder combination are subjected to varying displacements or forces. The granules or discs react to the movement by altering the spacing between adjacent particles; this changes the total amount of surface area, which is the dependent factor in determining the electrical resistance between particles or discs. The sensitivity factor, S, as defined previously, ranges from 30 to 150 for carbon devices (ref. 24).

The use of the carbon devices is limited because of their poor linearity, large hysteresis effects, and high temperature dependence. The response of a carbon pile, composed of carbon discs placed upon each other in a column, is approximately hyperbolic in shape when the resistance deviation is plotted against the displacement or force applied to the carbon pile. Hysteresis effects may be as large as 20 percent of the full range response. The temperature dependence of the devices is in the order of -3×10^{-4} to -10^{-3} per degree centigrade (ref. 24); that is, the resistance decreases about one part in a thousand for an increase in temperature of 1° C.

The carbon microphone uses a small button of carbon particles and is highly sensitive. However, its use as a calibrated instrument is limited.

IV. Variable Reactance Transducers

Variable reactance transducers use mechanical motion to change the a-c impedance of the transducer, providing linear displacement, force, or pressure measurements. The variable reactance may be inductive, as in a variable inductor or differential transformer, or it may be capacitive, as provided by a variable capacitor. In

either case the variation in a-c impedance also is a function of the frequency of the excitation voltage used with the devices. D-c excitation may not be used, as reactance refers only to a-c sensitivity.

A. Variable Inductance Transducers

The characteristics of an inductor are determined by its geometry. The number of turns of wire, the material of the core, or the shape of the windings could be varied to transform mechanical displacement into a changing electrical parameter. The change most usually made on a transducer is the material of the core of the inductor.

If ferromagnetic material (iron, steel, nickel) is used as the core of an inductor, the value of inductance achieved is relatively high. An air core, on the other hand, does not provide an appreciable inductive value. With a core of ferromagnetic material coupled directly to the moving member of the transducer, movement of the core in and out of the windings produces a considerable shift in inductance value. With the inductor connected into an impedance measuring circuit, such as a bridge circuit, the mechanical movement is transformed into an electrical output.

The value of reactance depends upon the excitation frequency and, to some degree, the value of the excitation voltage. If too high a current flows through the inductor, a saturation effect may diminish the sensitivity of the device.

B. Differential Transformers

A differential transformer is an arrangement of three inductors about a common core. When two or more coils share a common core, the coils are linked magnetically and are considered to be windings of a transformer, with the current in one coil inducing a voltage in the second coil. In a differential transformer, one coil, the primary winding, is excited by a source of alternating current, and the induced voltages are present in both secondary coils.

If both secondaries have the same number of windings, and the core of the differential transformer links both secondaries equally with the primary, voltages of equal amplitude appear across each secondary. The secondaries are connected in series-opposition, as shown in figure 11. The two secondary voltages are equal in amplitude, but opposite in phase, thereby canceling the secondary voltages, and a null voltage output is obtained from the series connection.

When the core is moved off-center, however, each secondary is not linked identically; one secondary produces a larger voltage than the other, and there is an output voltage. The more off-center the core is displaced, the larger is the differential output, and there is a good linear relationship between displacement and voltage amplitude over a wide range (see figure 12).

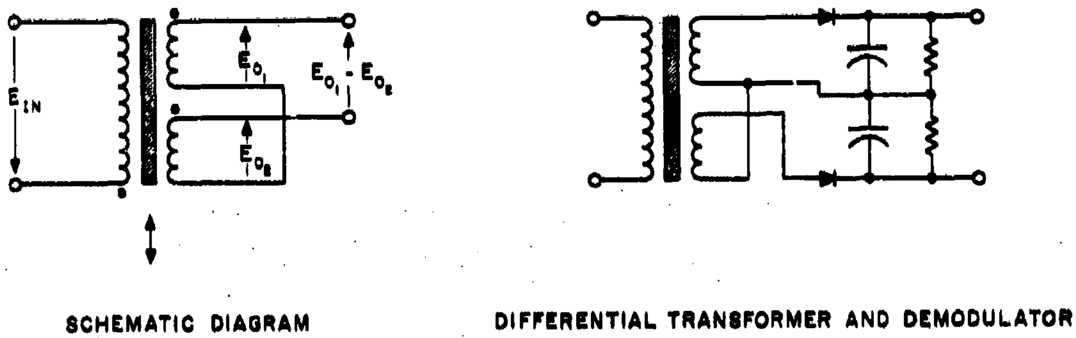


Figure 11. Differential Transformer, With Output Circuit

The output of the differential transformer may be converted easily to a d-c voltage, with the amplitude of the d-c voltage directly proportional to the displacement from the center position of the core and the d-c polarity a function of the direction of displacement. A full-wave rectifier and a filter section, as shown in figure 11, are used for this conversion.

C. Variable Capacitance Transducers

Variable capacitance transducers convert small capacitance changes to other measurable parameters, such as voltage and frequency.

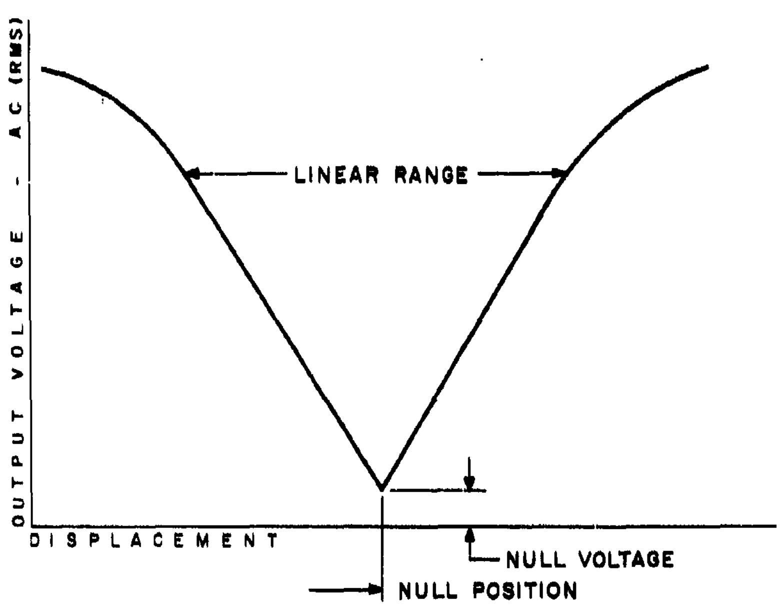


Figure 12. Output Voltage Characteristic of Differential Transformer

The generalized expression for the capacitance of a parallel-plate capacitor with a dielectric between the plates is

$$C = \frac{\epsilon A}{d}$$

where

C = the capacitance in farads
ϵ = the dielectric constant or permitivity in farads per meter
A = the area of the parallel surface of the plate in $meters^2$
d = the distance between the plates in meters

For a parallel-plate capacitor in air ($\epsilon = 8.85 \times 10^{-12}$), a parallel plate of 1 cm^2 ($10^{-4} m^2$) and a spacing of 10 mm ($10^{-2}m$) result in a capacitance of 8.85×10^{-14} farads or 0.08 picofarads. Changing any of the three factors, the permitivity, area, or spacing, changes the capacitance. Most variable capacitance devices change either the area or the spacing, seldom both simultaneously. The capacitance changes proportionally as the area of the plates, while varying the spacing changes the capacitance inversely.

A differential capacitor is essentially two capacitors arranged so that the moving element is the common capacitor plate for both. Figure 13 illustrates this arrangement. With this design, any movement of the center plate increases the capacitance in one capacitor while simultaneously decreasing the capacitance in the other.

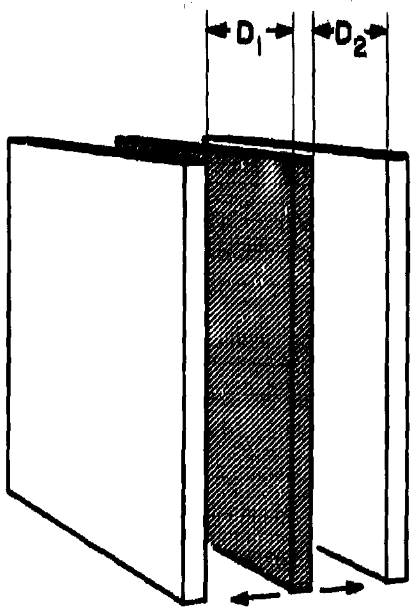


Figure 13. Functional Drawing of a Differential Capacitor

The capacitance variation does not have to be detected directly; i.e., the pressure fluctuations may be transmitted to the sensing capacitor plate through a catheter or tubing, as is done with piezoelectric and strain-gage devices. However, since the capacitor plates can be made from very thin, light metal, the frequency response of a system using the variable capacitor probably will be improved by having the mechanical movement displace the capacitor plate(s) directly.

Variable capacitance transducers have responses from less than one cycle per second to above the audio spectrum, a range that exceeds that required for physiological parameters. The range of sizes of the capacitive transducers varies, according to the specific application. The total volume required for the transducer itself may be less than one cubic inch. The mass of the devices is generally low, capable of being made less than an ounce.

Variable capacitance transducers are inherently continuous devices, affording infinite resolution. The sensitivity of these transducers is dependent upon the arrangement of the plates and the circuits associated with them.

One variation of a variable capacitor transducer* uses an air-coupled system to convert displacement variations to air pressures, which are then reconverted to variable displacements of a differential capacitor plate. The variation in capacitance is converted to a measurable voltage using Lion's ionization transducer (ref. 23), discussed in Section II on page 77.

TEMPERATURE TRANSDUCERS

In physiological monitoring, the temperature of the human subject provides a measure of the metabolic rate and well being of the subject under observation. In addition, temperature offsets can be used to observe other body parameters in a monitoring system.

I. Temperature-Dependent Resistance Devices

A. Theory

Practically every material, and especially the metals and semiconductors, exhibits a temperature-dependent resistance characteristic. In most metals, this is a positive coefficient of temperature; i.e., resistance increases with temperature. The reverse is generally true for semiconductors. The expression for resistance as a function of temperature is:

$$R = R_o \left[1 + \alpha \ (t - t_o) + \beta \ (t - t_o)^2 \right]$$

*The Decker Corp., Bala-Cynwyd, Pa.

where

R_o = the resistance of a specimen at temperature t_o
t = the temperature under measurement
α and β = coefficients of temperature resistance

Over a limited range of measurement, β is very small and the last term of the expression may be ignored. α is not completely constant and can vary over a temperature range. It is the constancy of α that determines the interchangeability and usefulness of a given device. Table III lists α at 20°C and 100°C for some metals.

TABLE III. TEMPERATURE COEFFICIENTS OF VARIOUS MATERIALS

Material	α $\left(\frac{1}{°C}\right)$	
	At 20°C	At 100°C
Aluminum	0.0039	0.0040
Copper	0.00392	0.0038
Gold	0.0034	0.0025
Manganin (84 Cu, 12 Mn, 4Ni)	±0.00002	-0.000042
Nickel	0.0047	0.0043
Platinum	0.003	
Silver	0.0038	0.0036
Tungsten	0.0045	

B. Resistance Thermometers - Description

A resistance thermometer or resistance temperature detector consists of a length of material, usually a metal, whose varying resistance is measured as an indication of the temperature to which the device is being exposed. The constancy and linearity of the device are dependent on the purity of the metal or metal alloy used. Platinum has the best overall linearity and, as a result, is used as the International Resistance Standard.

The resistance thermometer is composed of either a length of wire wound over a form, usually in a cylindrical shape, or a foil or grid of the temperature-sensitive material bonded on a flat backing in a similar manner as a strain gage. In either case, the resistance thermometer is the most accurate temperature sensing device, with an accuracy of $\pm 10^{-4}$°C at room temperatures.

When using the resistance detector, self-heating effects caused by current flow through the resistance wire must be avoided. No significant error is produced if less than 10 milliamperes is present in a wire of greater than 0.05 mm diameter (ref. 24).

C. Thermistors - Description

Thermistors are semiconductor temperature-sensitive resistance devices, usually having a negative temperature coefficient. The coefficient is as much as 14 times as large as that of metals in use. The resistivity dependence on temperature for semiconductor thermistors is of an exponential form:

$$\rho_t = \rho_o^{\frac{\beta}{t} - \frac{\beta}{t_o}}$$

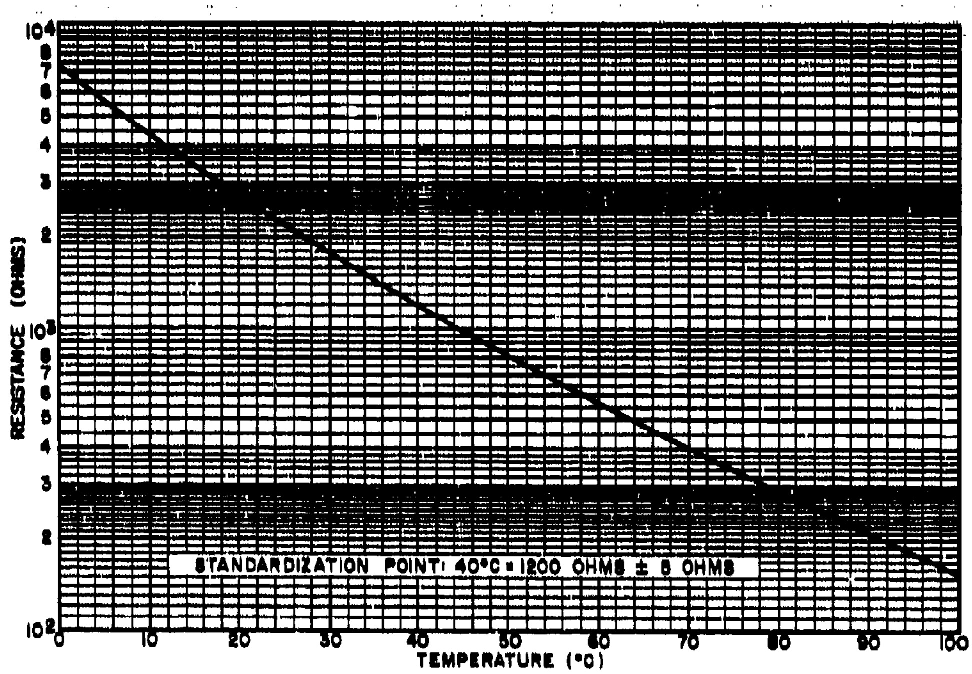
where

ρ_t = (ohm-centimeters) is the resistivity of the semiconductor at temperature t in °K

ρ_o = is the resistivity at temperature t_o, and

β = is a constant depending on the semiconductor content and processing (approx. 4000)

The value of β is unstable during the initial hours of excitation. Drift of over 1 percent may be expected. However, if the thermistor is aged, a stability of 0.2 percent per year can be achieved.



Yellow Springs Instrument Co., Inc., Yellow Springs, Ohio

Figure 14. Calibration Graph of Interchangeable Thermistors

The resistance range for a given thermistor can be specified according to the desired value for compatibility of equipment. Since the resistance of the thermistor varies exponentially with temperature, the overall resistance range for a temperature range of 100°C can be from thousands to hundreds of ohms. Figure 14 shows the characteristics of a thermistor series available from one manufacturer.*

Positive coefficient semiconductor thermistors also are manufactured. They are found useful for temperature compensation of negative temperature devices. One available device has a positive temperature coefficient of 0.7 percent per °C.**

Thermistors are made of molded powder and therefore may be shaped into any desired configuration. Small bead tips of 3/16 inch diameter are available as are flat disc types of twice this size. Figure 15 illustrates a few different thermistor configurations, all with the same resistance characteristics shown in figure 14.

The response time of a thermistor is highly dependent on the manner in which it is attached to the surface under investigation. Conduction of heat is usually the fastest manner of propagation; therefore, the thermistor must maintain a good thermal contact with the area in question. The tip of the thermistor probe containing the semiconductor material may be fabricated from steel, vinyl, glass, or some other material that does not react adversely with the monitored environment. The probe tip material and the total area of the tip determine the response time. The measure of response time is the time constant of the device, which is the time for the resistance of the thermistor to change to $(1 - 1/e)$ or 63 percent of the final resistance when the thermistor is subjected to a step change in temperature. Time constants for thermistors range from under 1 second to over 2 minutes. The time necessary for the instrument to respond to varying temperature conditions is determined by the time constant of the device.

Depending on the construction and attachment of the thermistor, the region monitored for temperature may be very small or many cubic feet.

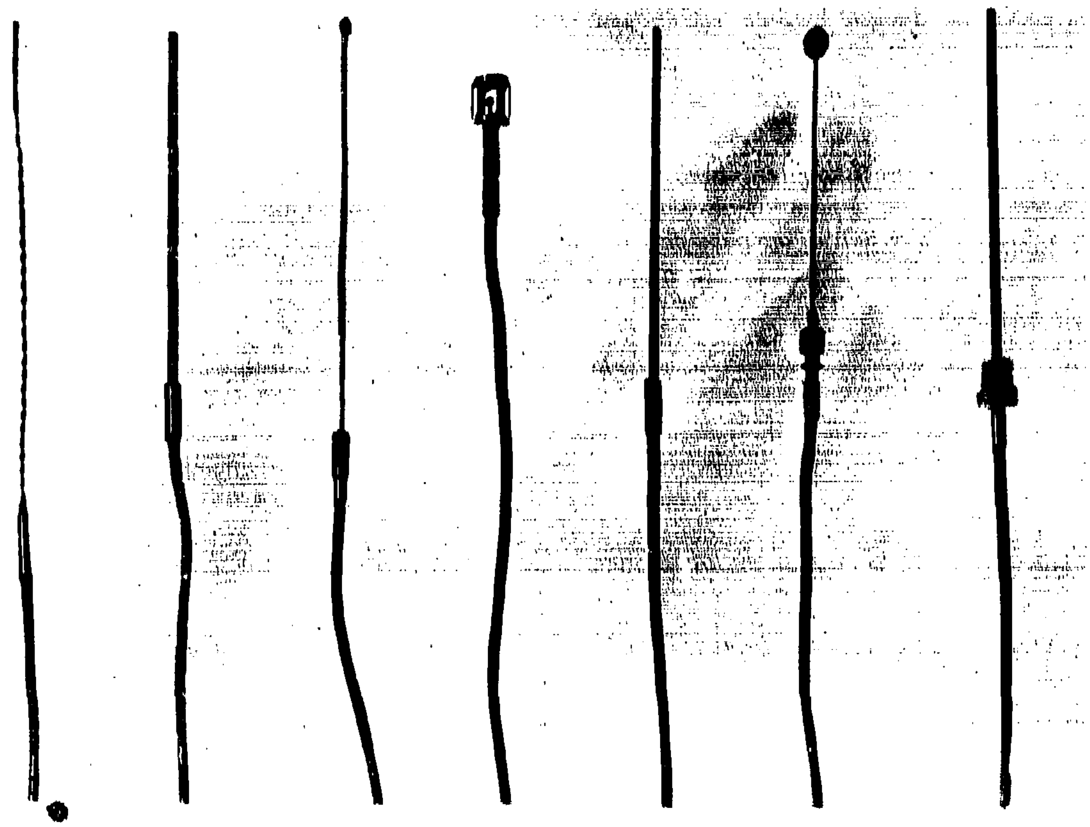
Since the tip of the thermistor is of some finite size, its contact with a point surface allows a portion of the probe to be in contact with the environment surrounding the point. The thermistor tends to average temperatures on its surface, restricting its application for point measurements.

II. Thermoelectric Devices - Thermocouples

A thermoelectric device generates an electrical signal proportional to the temperature to which it is exposed. The most common thermoelectric device is the thermocouple.

---

*Yellow Springs Instrument Co., Yellow Springs, Ohio
**Texas Instruments Inc., Dallas, Texas



Yellow Springs Instrument Co., Inc., Yellow Springs, Ohio

Figure 15. Some Typical Thermistor Probes

The principle of operation of the thermocouple is based on the Seebeck effect, which states that if a circuit contains two metals, with one junction of the two metals hotter than the other, a current flows in the circuit. The direction of the flow is dependent upon the relative thermoelectric coefficient of the metals and the relative temperatures of the two junctions. One junction is used as the measuring junction, and the other junction, referred to as the reference leg, usually is maintained at a constant temperature.

The standard metal combinations in use for thermocouple junctions are: platinum and platinum-rhodium, chromel and alumel, iron and constantan, copper and constantan, chromel and constantan, and carbon and silicon carbide.*

*Chromel (90% Ni, 10% Cr), Alumel (94% Ni, 2% Al, 3% Mn, 1% Si), Constantan (55% Cu, 45% Ni).

The expression for thermoelectric voltage is given by the equation:

$$E = At + \frac{1}{2}Bt^2 + \frac{1}{3}Ct^3$$

where

E is in microvolts,
A, B, and C are constants for a given metal in microvolts/°C, and
t is the temperature in °C

Thermocouple sensitivity, Q, is the change in voltage per degree C, or

$$Q = \frac{dE}{dt} = A + Bt + Ct^2$$

where A represents the thermocouple voltage at 0°C. A listing of sensitivity factors is given in Table IV. These sensitivities are given relative to platinum, thereby assuming that platinum is one of the junction metals. The sensitivity for any combination of listed metals may be determined by algebraically subtracting the two sensitivities from each other. The sign of the difference indicates the polarity of the voltage.

TABLE IV. THERMOCOUPLE SENSITIVITY OF VARIOUS MATERIALS (μVOLTS/°C)

Rhodium	6	Gold	6.5
Chromel	30.3	Iron	18.5
Constantan	-35	Nichrome	25
Nickel	-15	Germanium	300
Aluminum	3.5	Silicon	400
Silver	6.5	Selenium	900
Copper	6.5		

The voltages listed in the table are the open-circuit-voltage sensitivities obtained with the reference junction kept at 0°C. Since any measuring device must be connected to some indication device, there is the problem of loading. If current is drawn from the junction, heat is produced both in the wire and at the junction of the thermocouple, reducing the accuracy of the measurement. In addition, a voltage drop across the leads connecting the thermocouple to the indication device also interferes with accurate measurements. To avoid these potential causes of inaccuracies, a potentiometer circuit is used which draws no current in the balanced condition.

When using the thermocouple, the temperature of the reference leg of the device should be kept constant or, if there are variations in temperature, they should be known. Various schemes have been devised to compensate for temperature rises at the reference junction. One of these does so automatically through the use of a thermistor whose temperature characteristics are exactly the inverse of the thermocouple temperature variations.

III. Applications

Temperature-measuring devices are used in physiological monitoring to determine body temperature. Rectal detectors, oral detectors, and skin surface detectors are used and, generally, the rectal detector provides the most reliable results.

A rectal probe usually uses a thermistor with a beaded tip. The leads are long enough to allow the probe to penetrate three or four inches into the rectum, after which it is taped in place. Little discomfort is felt by the subject.

The oral detector is usually placed under the subject's tongue and held in place by either a harness or tape. This detector, however, often interferes with the normal activities of the subject being monitored.

The main limitation of the skin surface detector is the difficulty of maintaining good contact between the probe and the skin. One method has been to mount thermistors in the inner surface of a suit of underwear (ref. 37). The most popular method is to attach the thermistor to the skin with tape. Subcutaneous thermistor probes also are available, such as a hypodermic-needle-mounted probe or a probe intended for implantation in the subcutaneous tissue.

In addition to using temperature-detecting devices solely for determining body temperatures, thermistors and thermocouples are placed in the respiratory path of the subject to detect respiration patterns. As air is drawn into the mouth or nose and then exhaled, there are differences in air temperature. The output of the thermistor shows a decrease and increase of resistance corresponding to the inspired and expired air pattern. Figure 16 shows a record of respiration patterns obtained using a thermistor as the sensing element (ref. 27). The major disadvantage of this method of monitoring, however, is the need to position the transducer so that turning the head does not eliminate the monitoring function.

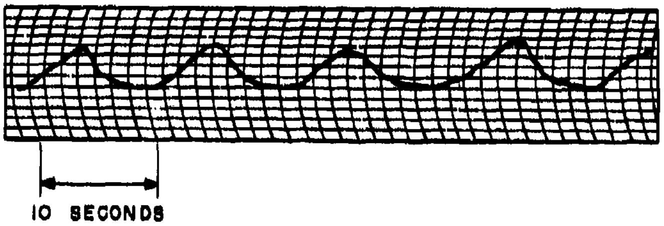


Figure 16. Respiration Pattern Obtained with Thermistor

OPTICAL TRANSDUCERS

Several physiological parameters can be measured with photoelectric devices. This is done by illuminating portions of the body and detecting the light transmitted through or reflected from the body tissue. Significant optical-to-electrical transductions are thereby obtained, since the body's optical characteristics will change with certain body functions, particularly blood flow.

There are three types of photoelectric transducers, each employing a different physical effect: photoconductive, photovoltaic, and photoemissive. These are all quantum effects, caused by the transfer of energy from the incident radiation or light to the current carriers in the sensitive material of the transducers.

All the photoelectric transducers discussed below can be operated effectively with common tungsten-filament lamps for their light source. Tungsten filaments emit light with a frequency spectrum that extends from the ultraviolet, through visible light, into the infrared. Figure 17 shows this spectral output. The devices under consideration all have peak responses to light of wavelengths within this spectrum. The spectral response is a particularly significant characteristic when selecting a device for operation in or near the infrared portion of the spectrum.

I. Photoconductive Transducers

A. Theory and Description

A photoconductive transducer is essentially a resistor that has the property of changing its resistivity upon exposure to radiant energy. Most photoconductors consist

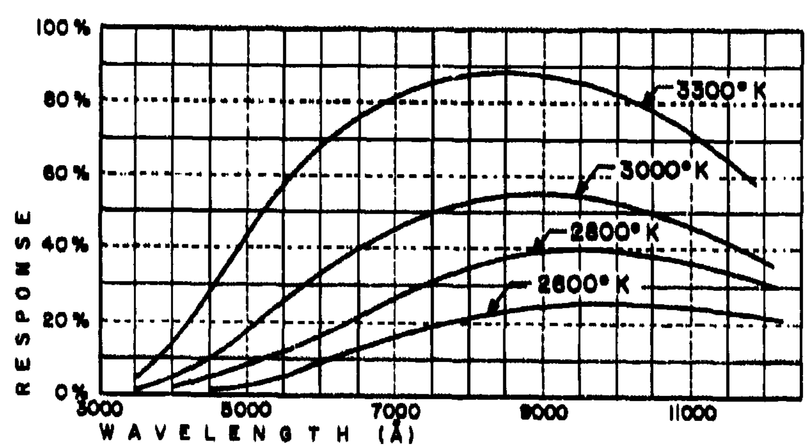


Figure 17. Spectral Output of Tungsten Filament Lamps

simply of a thin semitransparent film of resistive material deposited upon the inside of a glass chamber, which is then evacuated. Resistive materials most commonly used are selenium, the metal sulfides, germanium, and silicon.

The basic circuit for the photoconductive transducer is shown in figure 18. The physics of the semiconducting change in resistance is relatively complex. Simply put, when quanta of radiant energy strike the resistive material, giving up their energy, electrons are raised from the valence band to the conduction band in the material, and the resistance of the device drops sharply from its high, unilluminated level. With a voltage impressed across the device, the resultant increase in current flow is proportional to the amount of light striking the device. (The current is also a function of the amount of voltage applied.) Therefore, the response of photoconductors is essentially linear; that is, the drop in resistance and rise in photocurrent are proportional to the level of illumination. The resistance of the resistive material without illumination is in the megohm range, and it decreases to the kilohm range with 50 to 100 foot-candles illumination.

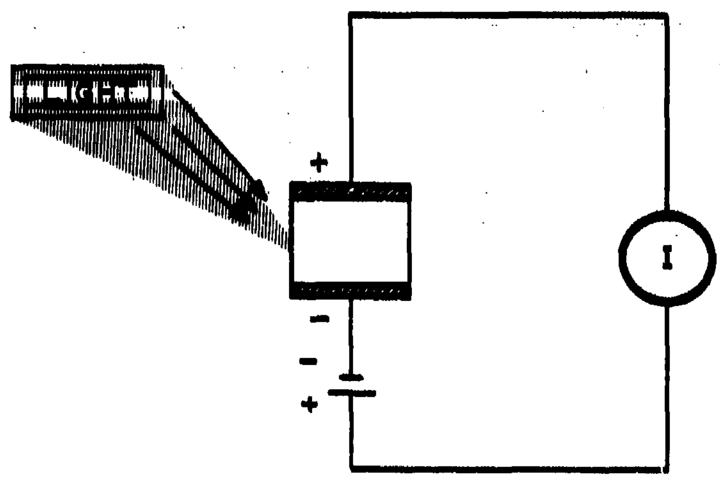


Figure 18. Basic Circuit of Photoconductive Transducer

The sensitivity of photoconductors varies widely with the resistive material, and is strongly affected by the voltage applied. Typically, for commercially available detectors using cadmium sulphide, the radiant sensitivity is about 300 to 500 microamperes per microwatt, at the peak wavelength of radiation. Spectral sensitivity may vary by several orders of magnitude, from milliamperes to several amperes per lumen.

The spectral response of photoconductive transducers matches the spectral output of tungsten lamps very well, with a peak response at about 500 millimicrons and a threshold response in the infrared range. The response in the infrared range can be

improved by cooling the cell (and by using special infrared-transparent glasses, such as sapphire or periklase).

The increase in current in a photoconductor is not instantaneous with exposure to light. The rise time is in the millisecond range, however, permitting photoconductors to be operated with lamps that are chopped or modulated at about 1000 cycles per second.

B. Applications of Photoconductive Devices

1. Pulse Monitor

Because of small size, sensitivity and availability in various current carrying sizes, photoconductive types are probably the most useful of all light-actuated transducers. However, their comparatively slow response to fluctuating light generally limits use to applications where frequencies above 1000 cps are not involved. In many applications either photoconductive or photovoltaic sensors may be used, the choice depending upon particular design considerations.

A photoconductive transducer is an excellent pickup for detecting the arterial pulse. The photoconductive pickup is placed over a convenient near-surface artery, and a 6-volt lamp adjacent to the transducer serves as a light source. The light from the lamp is reflected from the skin or scattered within the skin and is detected by the transducer. The intensity of the light detected varies as a function of skin opacity, which, in turn, changes with the arterial pulse. With the transducer connected into a bridge circuit, a changing output voltage, which indicates the pulse rate, is available for transmission, display, or recording.

This type of transducer is relatively easy and inexpensive to fabricate. The basic configuration is shown in figure 19. The transducer and lamp are anchored

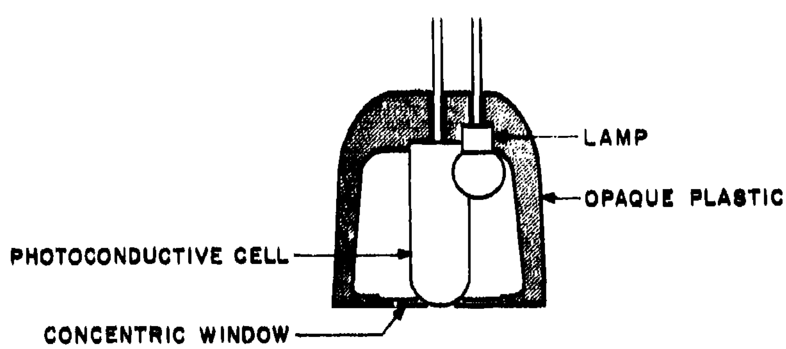


Figure 19. Photoelectric Transducer for Pulse Monitoring

permanently in a housing, which can be formed of an easily workable plastic, such as is used to make dentures. The plastic should be opaqued with black pigment to exclude ambient light. Several transducers of this type are commercially available.

2. Photoelectric Plethysmograph

The photoconductive transducer used to monitor the arterial pulse rate (described above) also can be used to monitor blood flow. The method of operation is essentially the same, except that the pulse waveform is modified and shaped to give an analog of the quantity (and, in some instances, the pressure) of the blood.

II. Photovoltaic Transducers

A. Theory and Description

The photovoltaic transducer is a semiconductor device which generates a voltage when exposed to light. It can be used, therefore, to measure light without an external, exciting voltage or current. Figure 20 shows the construction of such a device. Upon a base plate of metal (iron or steel), a layer of specially processed semiconducting crystal (usually selenium) is deposited. A thin, transparent metal layer is deposited on top of this, with an insulating barrier layer between. A metal collecting ring or electrode is usually then deposited upon the transparent layer.

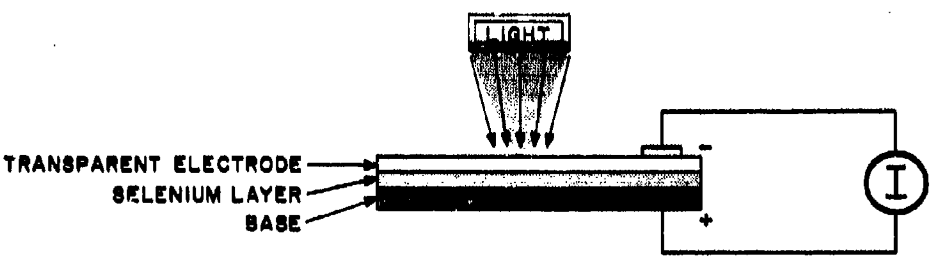


Figure 20. Construction of a Photovoltaic Transducer

When light falls upon the device, it penetrates the transparent layer and gives up its energy in the semiconducting layer, creating negative and positive charges across the barrier, which results in an electromotive force. Current will flow in an external circuit that is connected to the base and the collecting ring of a photovoltaic transducer.

Impedance matching with the photocell is difficult because of the relationship of output current and internal resistance. Figure 21 shows an equivalent circuit for a photocell. It is essentially a photocurrent generator, I_p, shunted by a capacitance

and an internal resistance, R_I, in series with a resistance, R_S. The current, I_R, through a load resistor, R_L, is expressed as

$$I_R = \frac{I_P R_I}{R_I + R_S + R_L}$$

The internal resistance, R_I, is extremely high in the dark or unilluminated photocell, but it falls rapidly with the level of light incident upon the cell. The rate of fall is proportional to the size of the load resistance. The most effective energy transfer occurs when the load resistance is small with respect to the internal resistance; however, it is difficult to keep the load resistance small under conditions of changing illumination.

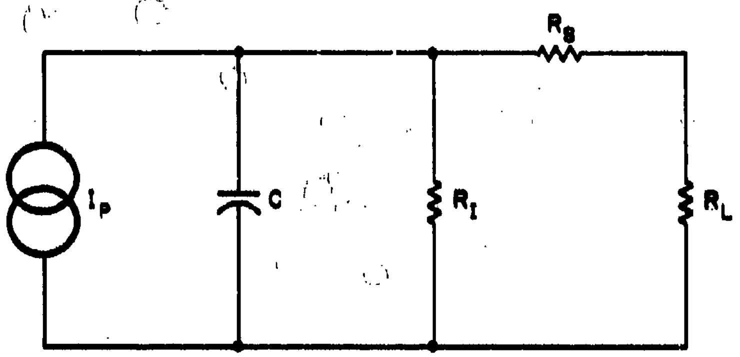


Figure 21. Equivalent Circuit for a Photovoltaic Transducer

However, in a typical physiological measurement, such as oximetry, with a constant light level from a tungsten source, the problem is minimized. In fact, changing the load resistance makes it possible for the user to adjust his instruments for either linear or logarithmic response. Thus, a simple change in load can effect a wide change in scale factor.

The open-circuit voltage of a photovoltaic transducer increases logarithmically with the level of illumination, and can go as high as 600 millivolts. The output current in a closed circuit, however, is essentially linear with illumination, except at very high illumination levels. Selenium photovoltaic cells have good sensitivity, on the order of 100 to 600 microamperes per lumen. Table V shows the typical output current obtained from one group of commercially available photocells with 100 foot-candles illumination.

The spectral response of photocells covers the visible light spectrum, and extends into the ultraviolet and the infrared. Selenium cells have a peak response at about 580 millimicrons. Cells of germanium and silicon have strong responses in the

infrared region (threshold response for germanium is about 1700 millimicrons; for silicon it is about 1200).

TABLE V. OUTPUT CURRENT OF TYPICAL SELENIUM PHOTOCELLS

Cell Type*	Dimensions (in.)		Photosensitive Area (in.^{2})	Output Current at 100 Foot-Candles, 100 ohms (μa)
	Diameter	Thickness		
A2	0.25	0.047	0.045	12
A3	0.38	0.047	0.06	20
A5	1.13	0.047	0.78	250
A7	1.50	0.058	1.40	440

*International Resistor Corporation

B. Applications of Photovoltaic Devices

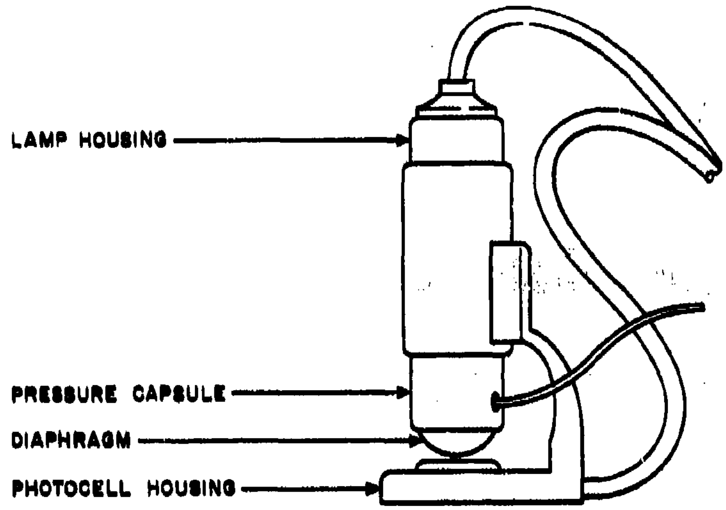
1. Oxygen Saturation

Photovoltaic transducers measure the percentage of oxygen saturation of the blood by detecting the difference in absorption by the blood of visible light and infrared light that have been passed through a portion of the body. The lobe of the ear is convenient for this measurement. (In clinical studies, samples of blood are withdrawn by catheterization and chambered for optical measurement in a device called the cuvette.) This measurement is called oximetry.

Figure 22 shows an earpiece designed for measuring oxygen saturation of the blood. The earpiece contains two selenium barrier-layer cells, each with a different optical filter. One passes only infrared light transmitted through the ear; the other passes only red. Either absolute or relative indications of the oxygen content are obtained, depending upon the design of the external circuits. To obtain a reference measurement of the light transmitted by the bloodless ear, a transparent, pneumatic pressure dome, located in the lamp housing between the lamp and the ear, is inflated to occlude the blood in the ear.

2. Blood Pressure

A fair approximation of systolic blood pressure can be obtained at the ear using essentially the same transducer as is used for oximetry. This relationship was uncovered when it was found that the pressure required to occlude blood flow completely through the portion of the ear lobe under observation corresponded to the systolic blood pressure (ref. 60).



The Waters Corp., Rochester, Minn.

Figure 22. Photoelectric Earpiece for Oximetry

For this blood pressure measurement, the air supply of the pneumatic dome in the earpiece is monitored with an appropriate pressure transducer, such as the strain gage. A comparison of the recordings of the pressure transducer and the photogalvanic transducer yields the systolic blood pressure. This type of measurement is not restricted to photovoltaic devices, of course. Photoconductive transducers could well be used. Photoemitters, however, are generally too bulky for application to this measurement.

III. Photoemissive Transducers

A. Vacuum Phototubes

The simplest photoemissive transducer is a vacuum tube containing a photosensitive cathode and a collecting anode. The anode is maintained positive by an external potential. When light falls on the cathode, electrons are liberated, and they move toward the positive anode, creating an electric current. Figure 23 shows a simple circuit for a photoemissive detector.

There are two fundamental relationships in the operation of this transducer. First, the current is directly proportional to the intensity of the light, and, second, the energy of the emitted electrons is related directly to the wavelength of the light, not the intensity.

Materials showing the greater emissivity occur in the lower order of the periodic table, where the total binding force upon electrons is lower. These materials include the alkali metals, particularly cesium. Photoemissive devices are classified by

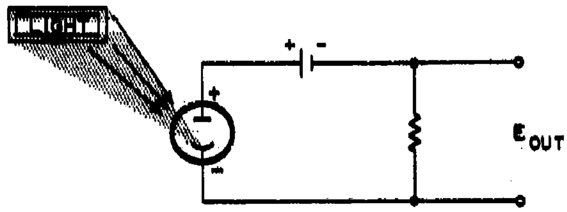


Figure 23. Basic Circuit for a Photoemissive Transducer

manufacturers according to their spectral sensitivity, which is a function of the material used for the cathode. The four classes of photoemitter of principal interest, and the photosensitive materials used in their cathodes, are the following:

S-1	Silver-oxygen-cesium
S-3	Silver-rubidium oxide-rubidium
S-4	Antimony-cesium
S-8	Bismuth-oxygen-silver-cesium or Bismuth-oxygen-antimony-cesium

The spectral sensitivity and relative response of these photocathodes are shown in figure 24. Three of the types show peak response at about 400 millimicrons, and threshold response falls off at infrared wavelengths. Type S-1, with an Ag-O-Cs cathode, and a peak response at about 840 millimicrons, has a threshold response which extends well into the infrared (about 1100 millimicrons). This characteristic makes the S-1 device useful in measurements where the spectrum of light detected is of greater significance than the quantity of light. If quartz rather than lime glass is used for the vacuum tube, threshold response out to the ultraviolet also is possible.

Significant performance characteristics of these devices are given in table VI. The following characteristics apply generally to all four classes:

1. Frequency response is uniformly high, in the megacycle range, since the only limit on photoemissive devices is electron transit time.

2. Output current generally should be limited to 1 microampere per square centimeter of cathode material, since higher current density causes unstable performance.

3. The current-voltage response of photoemitters is essentially linear. Above a positive anode voltage of about 60 volts, an increase in anode voltage

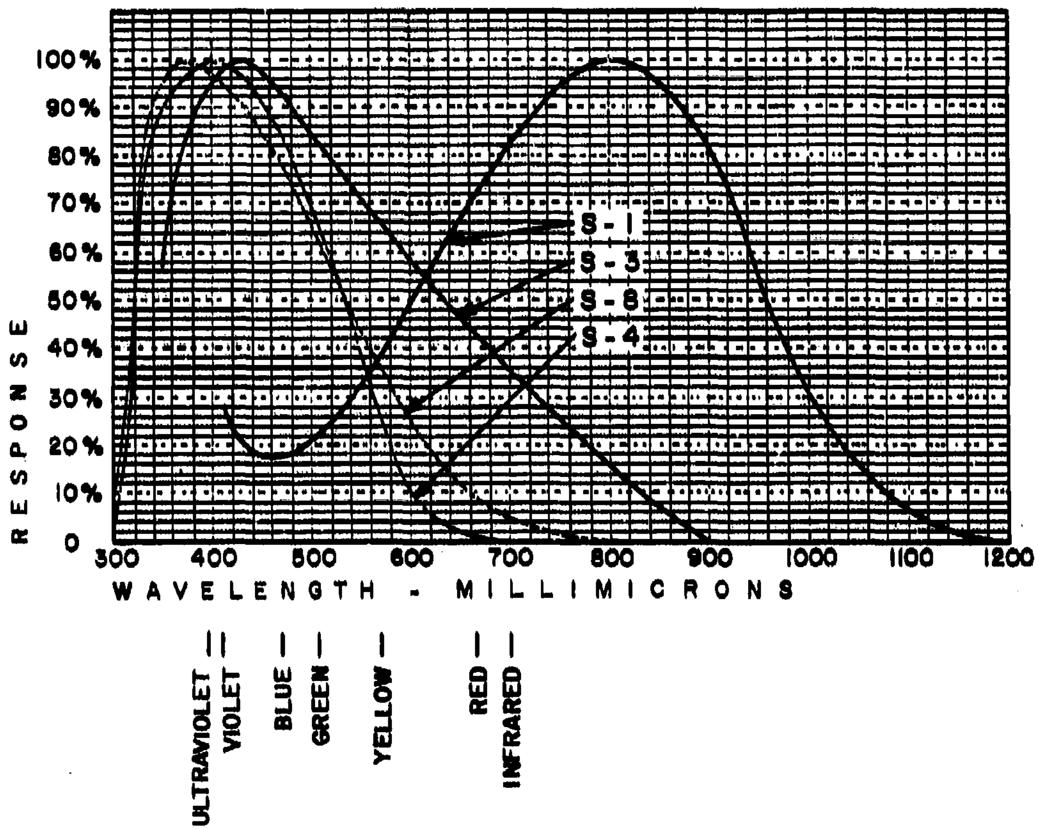


Figure 24. Spectral Sensitivity of Various Photocathodes

produces an insignificant rise in current. Most vacuum phototubes are operated with an anode voltage of +250 volts.

4. Output resistance of the devices is determined by the output or load resistor, which is usually between 1 and 20 megohms.

5. Dark current in vacuum tube photoemitters is about 5 to 10 microamperes, and can be minimized by cooling the cathode and lowering the applied voltage.

B. Gas-Filled Phototubes

Gas-filled phototubes are used instead of vacuum tubes where light input levels are low, and the output current obtainable with a vacuum tube is too low for accurate measurement. In the gas-filled tube, the collision of electrons (liberated from the cathode) with the gas molecules in the cell ionizes the gas, and there is a conse-

quent liberation of secondary electrons. By this means, as much as 10 times more current is obtained from the gas-filled tube than is obtained from the vacuum tube.

The operation of a gas-filled tube differs from that of the vacuum tube in several ways. For one thing, frequency response in the gas-filled tubes is considerably lower because of the increased transit time of the electrons necessitated by the ionization phenomenon. The upper limit on response of gas-filled tubes is about 10,000 cycles per second.

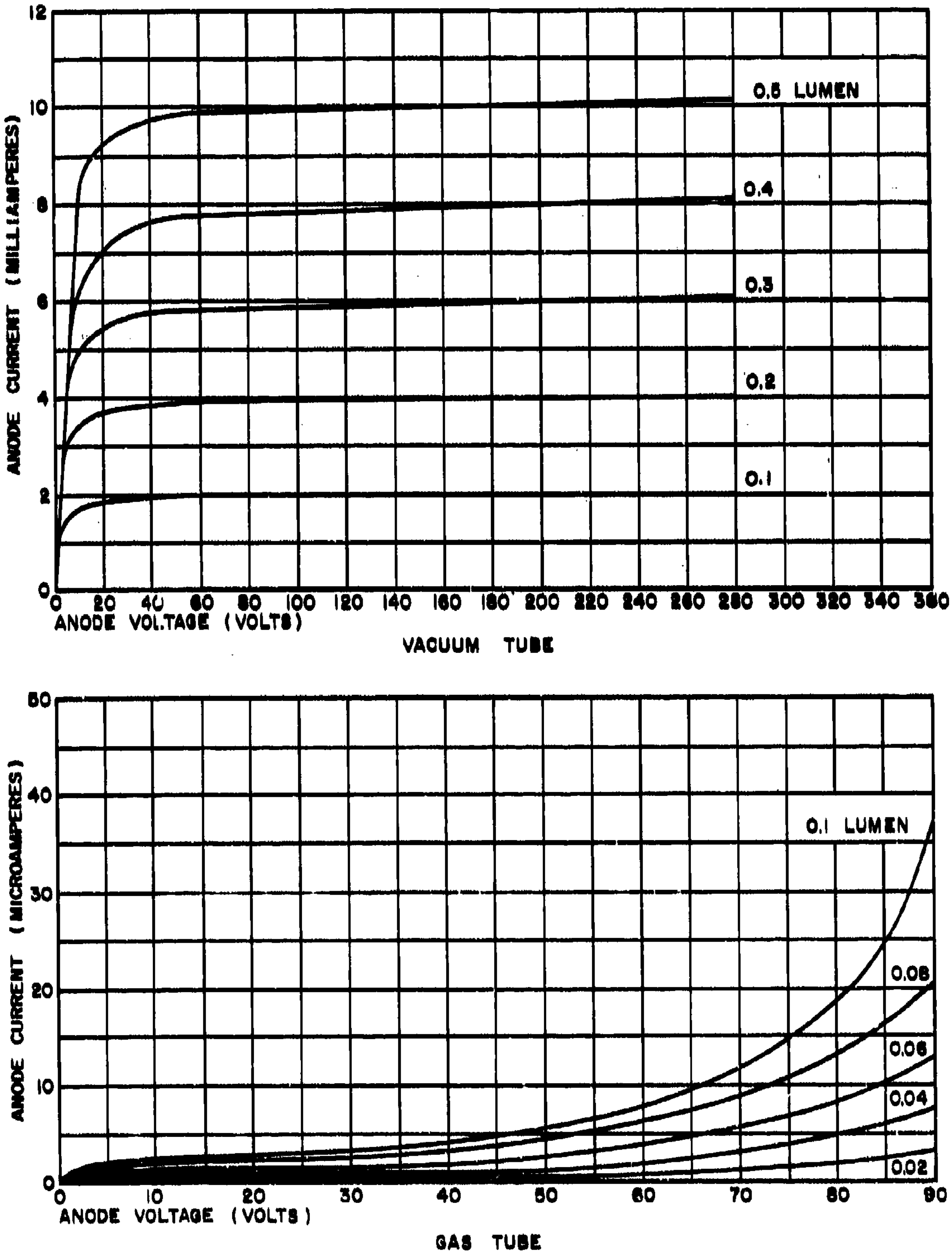
TABLE VI. CHARACTERISTICS OF TYPICAL VACUUM PHOTOTUBES

Class or Type	S-1	S-3	S-4	S-8
Spectral response				
Peak, millimicrons	840	420	400	400
Threshold, millimicrons	1100	900	700	800
Sensitivity*				
Luminous, microamperes/lumen	12-40	4-15	25-70	10-60
Radiant, microamperes/microwatt	0.0018	0.0016	0.042	
Dark current,* microamperes	0.005	0.005	0.0125	
Cathode current density, microamperes/in.^{2}	30	100	100	
Cathode current				
Average, microamperes	10	5	5	
Peak, microamperes		10	20	
Maximum temperature	100°C	100°C	75°C	100°C

*With +250 volts on the anode.

Also, in the gas-filled tube, the anode voltage-current response is nonlinear. Output current is dependent markedly upon anode voltage, and the plate voltage supply must be controlled more carefully than in a vacuum-tube circuit. Figure 25 shows the current-voltage relationship for the two types of phototubes.* The gas-filled tube must be used cautiously because an excess of anode voltage (normally, above 90 volts) can produce complete ionization and glow discharge, resulting in damage to the cathode. The gas phototube should be used for its greater sensitivity, but operated at lower ratings of anode voltage and current than vacuum phototubes.

*Radio Corporation of America, Harrison, N. J.



Radio Corporation of America, Harrison, N.J.

Figure 25. Current-Voltage Response of Phototubes (Vacuum and Gas)

C. Multiplier Phototubes

Another type of photoemissive phototube is the multiplier phototube. This device employs a series of secondary electrodes or dynodes, between the cathode and the anode, to amplify the cathode current internally. A positive potential of about 100 volts is maintained between electrodes. The tube uses a cascading operation, wherein electrons liberated from the cathode strike the first dynode, causing secondary emission. The combined electron flow then strikes the second dynode, creating even more secondary emissions. This process is repeated (through as many as 13 stages) until the combined electron flow from the last dynode is collected at the anode. A current gain of as much as 10^6 is possible (a luminous sensitivity for the cell of 20 amperes per lumen while that of the cathode is only 20 microamperes per lumen). Multiplier phototubes have obvious advantages, being by far the most sensitive of the phototubes, making possible usable outputs without further amplification. And gain is obtained without significant loss of frequency response, as in the gas phototube, because no ionization takes place, and the increase in electron transit time is negligible.

D. Applications of Photoemissive Devices

For many applications, the multiplier phototube appears to be the most attractive of photoemissive transducers, because of its extremely high sensitivity. However, the smallest tubes are between 2 and 3 inches long, and they require relatively complex, stable power supply circuits for efficient operation. Consequently, they are unsuitable for use in transducers that must be attached to the body directly, as, for example, the photogalvanic oximeter earpiece or the photoconductive plethysmographic pickup.

Multiplier phototubes, therefore, are usually restricted in use to clinical or laboratory measurements for various blood analyses, such as cuvette oximetry, blood cell counting, and gas absorption.

ELECTROCHEMICAL TRANSDUCERS

Chemical-to-electrical transduction is used for at least one physiological measurement: the concentration of oxygen, both in blood and in expired air. Polarographic transducers generally are employed for measuring oxygen concentration. In these transducers, current flow between two electrodes is a function of the amount of oxygen in the liquid or gas environment being measured.

1. The Polarographic Cell for Oxygen Sensing

The reference electrode in a polarographic cell is made of silver or a silver-silver chloride alloy. The active electrode generally is made of platinum (gold and calomel also have been used with success). The electrolyte usually is potassium chloride or sodium chloride. A small voltage (0.5 to 0.7 volt) is impressed across the electrodes,

causing a small current to flow. This current flow decreases as the cell becomes polarized by the concentration of hydrogen ions at the platinum cathode.

The electrodes and electrolyte are sealed behind a semipermeable membrane of polyethelene or Teflon. The membrane prevents contamination of the electrolyte when the cell is immersed in the fluid or gas being studied, but it permits oxygen molecules to pass through into the electrolyte. There, the oxygen molecules are drawn to the platinum electrode, combining with the hydrogen ions. This action depolarizes the cell, and there is an increase in current flow through the electrodes. The amount of current flow is proportional to the oxygen concentration of the gas or liquid adjacent to the membrane.

II. Applications

A. Oxygen Concentration in Blood

Oxygen concentration in the blood can be measured in two ways by polarographic means. In one, the polarographic cell is contained in a cuvette, and blood samples are taken from the subject by catheterization. In the other, the transducer is placed into the blood stream directly by incorporating the polarographic cell into the catheter. One miniaturized device is available that can be inserted into an 18-gage needle.* In this device, the active electrode is a 0.0005-inch platinum wire, which is sealed in glass with only the tip exposed. The wire is cemented into a fine silver tube that serves as the reference electrode. The tip of the electrode is covered with the polyethelene membrane.

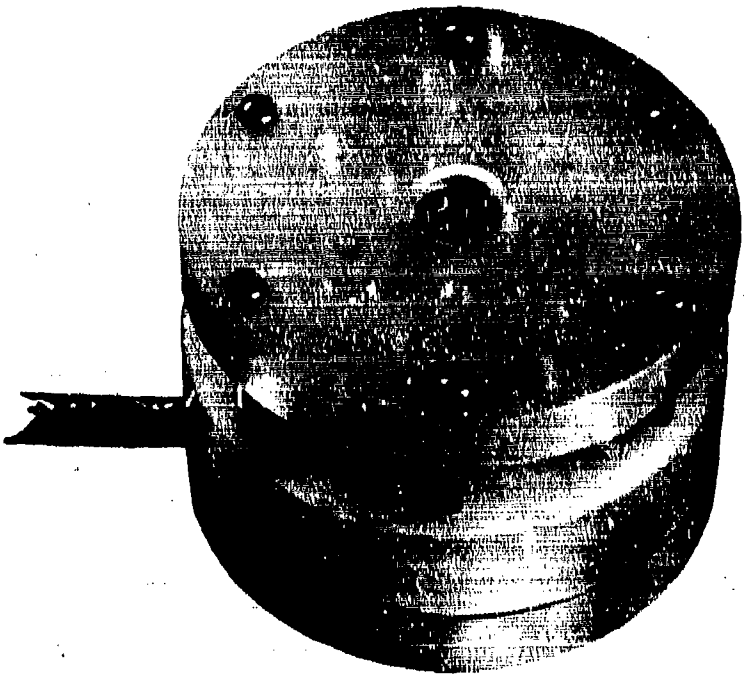
This method of measuring oxygen concentration in the blood is restricted to applications where venous or arterial puncture can be tolerated, and where adequate precaution against hemorrhage is taken.

B. Respiratory Oxygen

Polarographic transducers may be incorporated into gas lines for the measurement of oxygen in respired air. Some such devices are commercially available. One, shown in figure 26, is not strictly speaking a polarographic device. It uses no polarizing voltage, but generates its own voltage, and has an active electrode of gold and a reference electrode of cadmium (ref. 31).

Unlike polarographic cells immersed in the blood, those in a flowing gas stream are exposed to temperature changes. In some cases, changes in temperature can be compensated for by placing a thermistor across the electrodes and taking the output across the thermistor. The change in resistance in the thermistor varies inversely as the change in current caused by temperature, so the output tends to remain constant.

*Beckman Instruments, Inc., Spinco Division, Palo Alto, Calif.



Chemtronics, Inc., San Antonio, Texas

Figure 26. An Electrochemical Oxygen Sensor

Polarographic devices may be used in warning systems for hypoxia, which is the condition of low oxygen supply in the blood stream. For this purpose, the device is calibrated against a safe level of oxygen concentration, and a drop in electrical output below this level activates a signaling device. Even with the relatively slow response time of the device -- roughly 10 seconds -- it shows promise. However, the warning indication may or may not indicate a critical situation, since certain other parameters, such as ambient pressurization, the carbon-dioxide concentration in the expired air, and the rate of breathing, may affect the significance of the measured oxygen concentration.

OTHER TRANSDUCER CONSIDERATIONS

The sensing of physiological parameters, by the various means described above, yields information on body responses of interest to the physiologist or researcher. As part of an overall experiment or operation, however, other data inputs also may be necessary.

Most significant are data on the total environment (the conditions prevailing during the measurement). (Refer to Measurement Problems, Volume I.) To the greatest degree practical, the overall monitoring system should be planned with these environmental factors in mind. The following paragraphs therefore will discuss transducers for

environmental sensing, but only briefly, since such transducers properly belong in the realm of physical measurement, which is beyond the scope of this handbook on physiological measurement and monitoring.

Two other types of transducers are considered briefly also: those used for voice communication and for visual observation and recording.

I. Voice Communication

Often it is desirable to maintain voice communication with a subject in conjunction with monitoring his physiological reactions. One obvious reason, where abnormal and hostile environments are involved, is safety. The subject has, through voice communication, a means of signaling his desire to terminate, in an operation that may otherwise be out of his control. Further, in long-term exposure in hostile environments, voice communication may be vital to the subject's psychological well-being.

Voice communication also is important as a source of useful data. It makes available for correlation with measured quantitative data (1) an oral record of the subject's operating procedures, and (2) a report of his subjective observations.

The selection of a suitable microphone for voice communication is relatively simple. Good communications can be obtained within an audio bandpass ranging from 200 to 3000 cycles per second. These frequencies are well within the range of the most rudimentary microphones. The choice of microphone then hinges on two factors: background noise and location on (or near) the subject. If the environment during measurement includes a high level of background noise, a directional or contact microphone (throat or chest) may be required. The placing of the microphone on the subject will depend on whether the subject is wearing an oxygen mask, or whether he has other instrumentation attached near his throat and chest with which the microphone might interfere.

II. Visual Observation and Recording

Visual observation of the subject, as a backup for the measurement of quantitative data, provides a useful record of the circumstances and occurrences pertinent to the collection of data. Such visual records may be obtained continuously, with television or motion picture cameras, or selectively, with either of these or with still cameras.

If a visual record is all that is necessary, still or motion picture cameras can be used. Remote observation of the measurement while it is in progress obviously requires a television camera. Operation of the camera and the recording of its transmitted output may be continuous or intermittent, as desired.

A wide range of cameras is available, including numerous compact industrial type television cameras. Ambient light in the measurement scene probably would have to

be supplemented for successful observation: a flash unit with the still camera, or a suitable spot or flood lamp with television and motion picture cameras would be suitable. Photography or television pickup in color would require somewhat higher levels of illumination.

III. Related Environmental Monitoring

The environment to which a subject is exposed directly affects his physiological responses and condition. Thus, in many experiments, meaningful data must include quantitative environmental inputs. Atmospheric pressure and content are significant, as is ambient temperature. Also affecting the subject are factors related to the subject's movement through space -- acceleration and velocity. Ionizing radiation is a hostile environment that also should be monitored, both as a safety factor and to correlate long-term exposure with body responses.

The transducers used for displacement, pressure, and temperature are similar to those used for basic physiological monitoring; the major difference is the measurement range. The following paragraphs discuss briefly the monitoring of four parameters not previously covered: velocity, acceleration, atmospheric content, and radiation.*

A. Velocity Transducers

Most effects of velocity on the human subject are negligible when compared to the effects of change of velocity (acceleration). Under special conditions, however, the rate at which various support equipments operate, such as the speed of shafts, vents, and valves, or the rate at which certain liquids flow, either in the environment or in a cooling system, may be of interest.

To measure velocity, the linear motion of the body under consideration is compared to some other fixed or moving system, and the result is the relative velocity between the two systems. The transducers used most often are electromagnetic. Linear velocity transduction is obtained by the motion of a permanent magnet relative to a coil of wire. This type of transducer is used only for comparatively short-stroke linear displacements (0.5 to 5 inches). Sensitivity is on the order of 150 millivolts/inch/second. Resolution is infinite, as the device is continuous in operation. It requires no external power source. Accuracies of better than 1 percent are available. Analogous, transducing devices are available to measure the angular velocity of a system. In this application, the transducer is essentially a tachometer.

A flowmeter system of detecting the velocity of a fluid also uses the electromagnetic principle. In this type of transducer, the fluid is a conductor whose velocity or flow rate is of interest. A magnetic field is generated at right angles to the path of

*For more detailed coverage, see references 24 and 34.

the liquid in a pipe or tubing, and metal electrodes insulated from the pipe wall penetrate the wall. The voltage induced between the electrodes is proportional to the fluid velocity and the strength of the magnetic field.

B. Acceleration Transducers

Acceleration, or the change in velocity with respect to time, can greatly affect the subject's physiological and psychological responses. In space flight applications particularly, precautions are taken to guard the body from the distressing and destructive g forces of expected accelerations.

The accelerometer may be positioned directly on the subject to measure the exact acceleration to which he is being subjected. In most applications, it suffices to have the transducer in the same general area of the subject.

The accelerometer also is used to measure vibration and shock. In the same manner that straight accelerations are measured, the device measures the periodic accelerations due to vibratory movements, or the sudden but brief accelerations caused by shock situations.

Acceleration is measured easily because of the inertial force associated with changes in velocity (Newton's first and second laws of motion). This inertial force causes a known mass to move or exert a force on a calibrated electrical force detector, providing an electrical signal proportional to the acceleration. The force detector may be a piezoelectric device, a differential transformer, a strain gage, or a variable reluctance sensor. The total arrangement, including the known mass and the force detecting apparatus, is called the accelerometer.

Ranges available for accelerometers are from less than 1 g to over 500 g. (g is the acceleration equivalent to the acceleration produced on a given body by the force due to gravity, 32.2 feet per second per second at sea level.) If piezoelectric or potentiometric devices with composition or slide wire transducers are used, infinite resolution is provided because of the continuous output. Some accelerometers use a wire-wound resistor to provide a variable resistance output, somewhat limiting the resolution.

The accelerometer output depends on the type of transducer used internally. Resistance ranges of 1000 ohms, 10,000 ohms, and greater are available. Some devices require excitation and have an output of from 0 to 6 volts dc (used for telemetering purposes).

C. Transducers for Environmental Gas Analysis

Environmental gas transducers are used for the examination of oxygen or carbon dioxide concentrations in closed system determinations. The oxygen detector may

be used for a hypoxia warning device. The effect of artificially composed atmospheres on the physiology of a subject also may be of some interest, and the various gas transducers provide a means for monitoring this parameter.

One of these devices, used for carbon dioxide measurement, is shown in figure 27. The two infrared sources are modulated by the chopper fan. One beam passes through a reference cell into the detection chamber, while the other goes through a sample cell before entering the detection chamber. The detection chamber (composed of two-pressure-tight compartments separated by a diaphragm) and the sample cell are both filled with the gas under investigation. More light is absorbed by the upper chamber and sample cell than by the lower chamber alone, causing lower pressure in the upper half of the detection chamber. The diaphragm separating the detection chamber reacts to the pressure differential by moving toward the metal plate suspended in the chamber, changing the electrical characteristics of the transducer output.

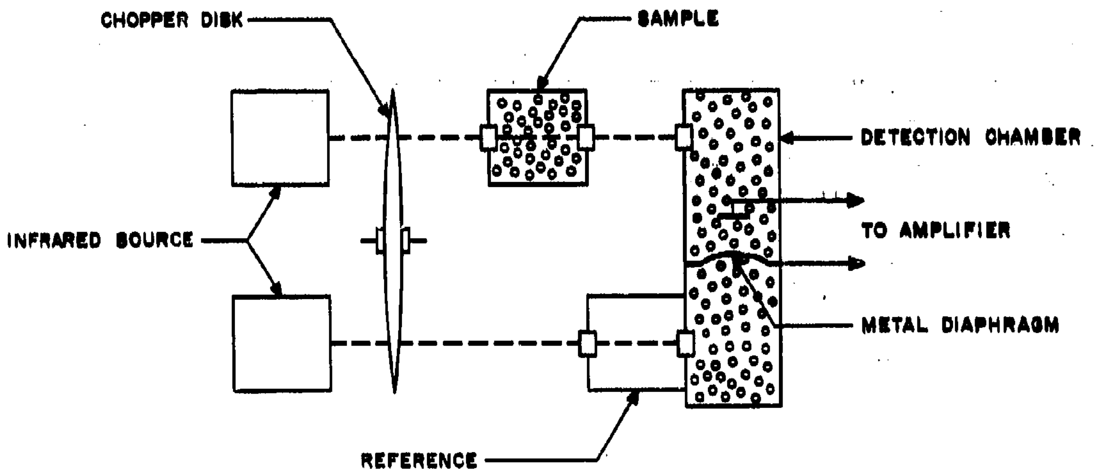


Figure 27. Carbon Dioxide Gas Analyzer

The response time of this device is 0.1 second for 90 percent indication of the final output. When used with associated control equipment a sensitivity of as low as 1 percent CO_2 concentration is available, with an accuracy of 0.03 percent.

The device also may be used to measure concentrations of other gases. Polarographic devices also are available for environmental oxygen monitoring.

D. Radiation Detectors

If there is any possibility of a radiation hazard in an environment where physiological monitoring is being performed, immediate knowledge of its presence and

degree are required, because radiation is a parameter that does not immediately react on or show up in the physiological responses of the subject, except for extremely high doses.

The three most likely types of radiation that may be encountered are alpha particles, beta particles, and gamma rays. Of the three, the gamma radiation usually is the most dangerous because of its higher energy and greater penetrating power.

The types of radiation detectors adaptable for instantaneous readout are either the sealed electrode type, in which a current is produced because of the ionization of a gas between two electrodes, or a photosensitive device that reacts to visible radiation produced by nuclear radiation in certain media.

Gas-filled detectors are of three varieties, all based on the same mechanical configuration. The main differences in their characteristics are the pressure of the gas enclosed in the transducer and the voltage impressed across the electrodes. The ionization chamber detects each individual ionization event produced by a collected particle or ray and produces a very small current in direct proportion to the number and strength of the impressed radiation effect. The proportional chamber relies on an ionization multiplication effect to produce many ionizations from each one detected, creating a larger current than is available from the ionization chamber, for a given amount of radiation, and thus requiring less amplification to drive a display or recording system. When the electrode voltage is great enough, the gas filled detector produces a large output pulse of constant amplitude for any particle detected, irrespective of the particle energy. In this operation, the tube is referred to as a Geiger-Müller counter and it is an extremely sensitive particle quantity detector; it is not useful, however, for determining the energy of each particle detected. The time required for the Geiger-Müller counter to recover, after each particle detected, is relatively long compared to the other two modes of operation of the gas-filled detector, limiting its use in radiation fields where particles are arriving at rates faster than every 200 microseconds.

A variation of the gas-filled detector is the crystal detector, in which the electrodes are separated by a crystal such as silver chloride, zinc sulphide, diamond, cadmium sulphide, or one of the thallium halides. The operation of the crystal detector is similar to the gas-filled device, with faster counting speeds obtainable because of the shorter time required for an ionization effect to travel in the crystal lattice than in the molecular structure of the gas. However, crystal detectors become polarized eventually, requiring replacement after long exposures to radiation.

Radiation detectors using a photo effect are the scintillation counter and the Cerenkov counter. The scintillation counter operates on the principle that certain crystals exhibit fluorescent properties when exposed to radiation. Crystals of organic structure (anthracene, stilbene, naphthalene), inorganic composure (ZnS(Ag)), or artificial plastics (polystyrene or polyvinyl toluene +3 percent terphenyl +0.02 percent

tetraphenyl butadiene) are cemented to the face of a photomultiplier that detects the fluorescence produced and generates a large electrical output pulse for each particle entrapped in the crystal, with the pulse amplitude proportional to the energy of the particle.

The Cerenkov counter is similar in construction to the scintillation counter, using a crystal of high dielectric constant like polystyrene or lucite on a photomultiplier face. This material emits visible light when a high-velocity charged particle enters its confines.

Section II

SIGNAL MODIFIERS

GENERAL

The electrical signal produced by a transducer responding to a physiological process in the body generally is not in a form suitable for display or interpretation. Usually the signal is not only weak, containing insufficient power in itself to activate meters or recording devices, but frequently is accompanied by extraneous information or artifacts, detracting from or masking the desired information. Signal modifiers, such as amplifiers, differentiators, integrators, and filters, are used to modify, process, and normalize this signal.

The location of a signal modifier in a physiological monitoring system depends largely on the individual system. Generally, signal modifiers immediately follow the transducer and precede a stage of transmission or recording.

Signal modifiers are manufactured by many different suppliers. Each supplier, and indeed each model that each supplier manufactures, has its own specifications, with differences ranging from package size to frequency and amplitude capabilities. By understanding the meaning of the various modifier specifications, the proper equipment can be obtained or, if necessary, constructed so that the desired physiological function is monitored successfully.

AMPLIFIERS

1. General

The basic function of an amplifier is to increase the voltage, current, or power (work-performing capability) of an electrical signal. The electrical signals representing most biological phenomena generally are very small, whether obtained by electrodes or transducers. Amplifiers increase these electrical signals so that more varied and dependable recorders can be used, and biological signals can be transmitted over long distances. This simplifies the problem of accumulating time correlated data.

Power amplification is achieved by using the original small biological signal to control the flow of power in a secondary circuit, with the secondary power source independent of the input signal. Sometimes other modification processes, such as filtering or limiting, are performed simultaneously with amplification. These other functions, however, are considered separately in the following text as independent actions

II. Amplifier Characteristics

An amplifier, or amplifier stage, is described by its response and characteristics as related to the input signal and the amplified output level. Amplifiers are compared and selected for an application according to the combination of these parameters (the amplifier characteristics).

A. Gain

One of the major parameters of interest in discussing amplifier performance is the gain of the amplifier stage. Voltage gain is defined as the ratio of the output signal level with respect to the input level. In most situations the voltage gain is the prime consideration, since most physiological signals are in the microvolt (1/1,000,000 of a volt) to millivolt (1/1000 of a volt) range, and most display and recording devices require input voltages of a considerably higher level. The gain obtainable from an electron device, either vacuum tube or transistor, is determined by the characteristics of the device, and the manner in which it is connected in the amplifier circuit.

Because gain is a ratio of like quantities, it is dimensionless. When stages of amplification are connected in sequence, the total gain is the product of the gains of the individual stages. For convenience another manner of expressing voltage gain, the decibel, has been derived. The decibel (db) is based on the logarithm, and is defined as follows:

$$db = 20 \log_{10} \frac{V_o}{V_i}$$

An example of the conversion between absolute gain and db's follows:

If an amplifier stage has a voltage gain of 100 (the output voltage V_o is 100 times the input voltage V_i), the gain in decibels is:

$$\begin{aligned} db &= 20 \log_{10} 100 \\ &= (20) \quad (2) \\ &= 40 \end{aligned}$$

or there is a gain of 40 db. If a following stage also has a gain of 100, the total absolute gain is (100) (100) or 10,000. In the decibel notation, the gains of the two stages are added, and the result is 80 db gain. Both answers indicate identical voltage gains, which fact is confirmed by substituting the db value in the formula:

$\log_{10}$ 10,000 is 4, giving a total gain of (20) (4) or 80 db.

Besides the convenience of being able to add decibels instead of having to multiply absolute gains, the decibel notation also is useful in noting power variations. For power considerations, the decibel definition is:

$$db = 10 \log_{10} \frac{P_o}{P_i}$$

If a signal decreases in power output by one half, the corresponding decrease in power ratio is about 3 db. Thus for every 3-db loss in a system, the power content of the signal is halved.

B. Frequency Response

The amplitude versus frequency response of an amplifier is the measure of its capabilities to provide the same gain, no matter what the frequency of the input driving signal. Physiological signals frequently require amplifiers having responses from 0.01 to 5000 cycles per second. Generally, the frequency limits necessary for monitoring physiological phenomena are not demanding, and neither vacuum tubes nor transistors create any amplitude or phase frequency-response problems.

The phase relationships between the input and output signal may change with frequency. Phase shift problems do not occur at low frequencies, in simple direct-coupled amplifiers. However, in resistance-capacitance-coupled amplifiers, phase shift is inherent. A complete specification of the phase characteristics of an amplifier usually includes the entire frequency band under consideration. At the midrange frequencies, usually little phase shift is encountered, or if phase shift does occur, it is constant throughout this region. At the low and high ends of the band, however, the phase relationship between the input and output signals of an a-c amplifier may change rapidly with frequency variation. Figure 28 is a plot of the phase-frequency characteristics of a typical audio range amplifier.

Phase shift occurs in every stage of amplification where reactive components are used. Effects are generally additive, increasing with the number of stages. In

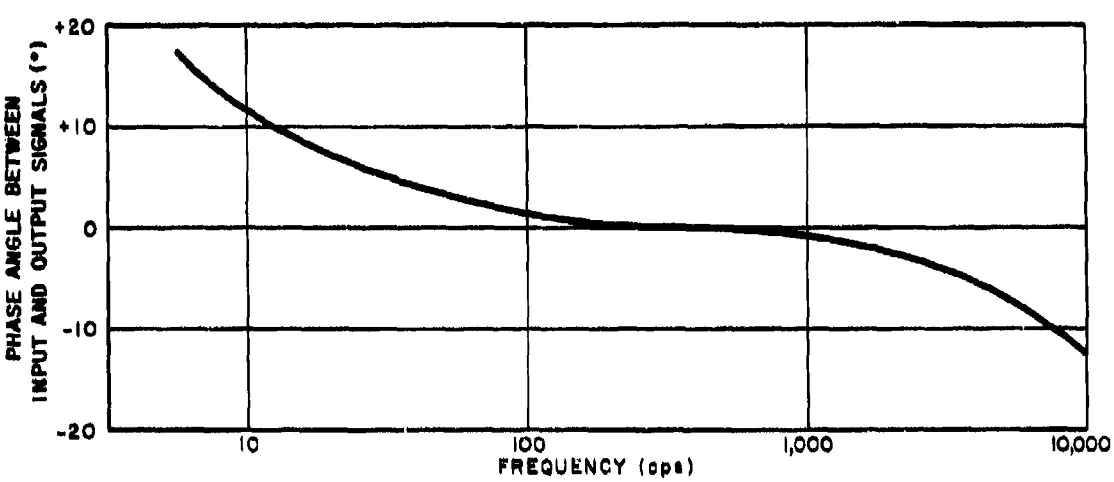


Figure 28. Phase Versus Frequency Characteristic of an Amplifier

amplifier designs employing feedback, certain combinations of phase and gain can lead to oscillation and complete blocking of usual amplifying functions.

If the phase shift characteristic becomes significant and threatens the usability of the amplified signal, special corrective phase-shifting networks may be included in the amplifier to compensate for the shift in the regions of interest.

C. Impedance (Input and Output)

The input impedance is the ratio of the voltage applied to the input to the current drawn. In most measurement applications, it is desirable to have the input impedance of the first stage of amplification as high as possible. In this way, a voltage impressed across the input stage draws very little current from the source transducer; consequently, little power is required of the source itself.

Output impedance may be defined as the ratio of amplifier open-circuit output voltage with respect to short-circuit output current. In most applications, if maximum power transfer is not a consideration, low output impedance is generally favored; impedance in the following stages is less a problem and better uniform response of the amplifier is maintained for varying frequencies.

If additional circuit impedances are placed in series with the input and output leads of the various amplifier configurations, these additional impedances must be considered as adding to the inherent impedance of the stage.

D. Drift and Stability

For an amplifier to be of value as a calibrated measuring instrument, it must be stable in its operation, with reproducible results for identical measurements. In d-c amplifiers lack of long-term baseline steadiness is termed drift, i.e., a slow change in output without a corresponding change in input. Drift generally arises in the early stages of an amplifier from (1) changes in supply voltages and (2) changes in the characteristics of the circuit elements with temperature or time. These latter changes occur primarily in active elements such as vacuum tubes or transistors, with temperature-sensitive resistors a secondary source of instability.

The effects of drift arising from power supply variations may be combatted in two ways. The most direct way is to stabilize the d-c supply voltage. However, when a power supply is not to vary more than a microvolt for a supply voltage of 300 volts dc, better than a millionth of a percent regulation of the power supply would be required, a rather stringent and unrealistic requirement. A more practical method of achieving stabilization is to isolate each stage of amplification; by using a resistor-capacitor network for interstage coupling, steady potentials are not passed, and slight shifts in the bias supply are not transmitted and amplified. However, the low frequency or d-c response of the amplifier must be sacrificed to obtain this stability, and the

limitations in amplitude frequency response, and phase frequency response must be recognized and accepted.

Drift arising from changes in circuit element characteristics is minimized by (1) care in selection of components for stability, (2) compensating circuitry, and (3) mechanical design or arrangement for minimizing temperature changes in the critical elements.

Another type of instability which may be encountered is due to undesired interstage couplings. Most commonly the source of the undesired coupling is the power supply. This type of instability is evidenced by oscillations that are of very low frequency (motorboating) or more rarely very high (ringing). Unless these unstable conditions are detected and removed, the oscillatory waveform mixes with or overwhelms the actual biological signal and produces an erroneous and probably useless output.

E. Hum and Noise

Hum and noise are extraneous signals which are superimposed upon the biological signal in the amplifier. Hum usually is generated as inductive or capacitive pickup of the power-line frequencies. It also may enter the signal path as a-c components in the output of the amplifier power supply. All other extraneous signals are considered to be noise.

Noise maybe generated from nearby ignition, motor, or transmitting equipment and radiated or conducted into the biological measuring system. In addition, noise is produced by the action of electron flow in vacuum tubes (shot noise), resistors (thermal noise), and transistors.

The usual measure of noise in an amplifier is the signal-to-noise ratio, expressed in decibels. In this calculation, the noise output of the amplifier is figured as the power output of the amplifier without signal input divided by the maximum power output capability of an amplifier, and is expressed in db as a ratio. This is a noise figure.

Because the input stages of an amplifier chain operate at lower signal levels, most of the noise picked up in an amplifier can be traced to the input stages. To minimize the input noise, the input tubes and transistors are carefully shielded, as is the input cable to them. If a vacuum tube is the input stage, the filament often is energized with dc rather than ac to minimize hum pickup. Amplifiers operating with very small signals sometimes use battery supplies for both filament and plate potentials to eliminate any hum interference from power lines.

F. Transient Blocking

When stages of amplification are coupled by resistance-capacitance circuits,

a brief signal or noise pulse in the input, having a higher than normal signal amplitude, can "paralyze" the amplifier temporarily. In some physiological amplifiers with an extended low-frequency response this blocking may last for many seconds. The time required to recover normal conditions after application of a specified surge type signal is called the recovery time. This recovery time may be ten times as long as the time constant of the coupling circuit.

G. Common Mode Rejection

Common mode rejection is the ability of a differential amplifier to ignore a signal that is present simultaneously at both input terminals of the amplifier. The measure of this ability usually is expressed in decibels, comparing the amplitude of the common signal (undesired) at the amplifier output to the difference signal (desired) at the output. Thus a common mode rejection figure of 60 db means that the desired differential signal is 1000 times as great as the common mode signal at the output of the differential amplifier (60 db = 20 log 1000).

The major drawback of a single-ended input amplifier in physiological work is the lack of common mode rejection. Noise picked up in common mode sources is indistinguishable from the desired physiological signal and is amplified along with the desired signal. (Refer to the discussion of differential amplifiers on page 71.)

H. Linearity

An amplifier operates linearly if the output signal is exactly proportional to the input signal according to a ratio which is the gain of the amplifier. For truly linear operation, the output-input relationship must hold for all levels and for all frequencies. The deviation from this ideal is one measure of the linearity of the device.

Nonlinearity generally is caused by improper operation of vacuum tubes and transistors. Under some circumstances it is introduced in transformers or other ferromagnetic devices. Nonlinear distortion produced upon sinusoidal signals is termed harmonic distortion. The quantity of harmonic frequencies produced is another measure of linearity.

Negative feedback, and push-pull and differential amplifier configurations are frequently used to minimize nonlinearity.

I. Power Output

Next to gain probably the most important overall characteristic of an amplifying system is its capability to deliver signal power to an external load. This factor is most commonly expressed as a figure in watts at a frequency within the working range that can be supplied to a resistive load without exceeding some definite distortion figure. For instance, 10 watts at 400 cps not exceeding 2 percent total harmonic

distortion would be typical for an audio-frequency amplifier. Because of the difficulties encountered in measuring harmonic distortion at frequencies below 60 cps, the output capability of physiological amplifiers is sometimes defined as peak-to-peak current or voltage delivered to a resistive load without departing more than 1 percent from linearity.

J. Dynamic Range

Dynamic range describes the capability of an amplifier to handle a range of signal amplitudes or power. For large signals, the range is limited by linearity considerations, i.e., signal distortions. For small signals, the range is limited by noise and stability factors. The ratio of the maximum to the minimum signal that the system will handle (while maintaining stated conditions of signal integrity or accuracy) is the dynamic range. Dynamic range may be expressed as a ratio or in db. It is always less than the signal-to-noise ratio.

III. Classification

A. The Stage Concept

The key components in the amplification process are the active circuit elements, such as vacuum tubes and transistors. Combined with the necessary passive elements, one or more of these active elements constitutes a stage of amplification. As noted before, the amplification process is the utilization of a small signal to control a greater amount of power, voltage, or current in a secondary circuit. The secondary power source may be an electrical battery or other energy source, generally termed the power supply.

1. Active Elements

The active element in a stage of amplification may be any of several control devices, although vacuum tubes and transistors are used primarily in physiological monitoring. Linearity of amplification depends on the degree of accuracy with which the variations in the output follow the signal variation to the input of the device. Vacuum tubes and transistors essentially are nonlinear devices; however, each has a quasi-linear region, within which they may be used. The active elements have little effect on frequency characteristics in the range of physiological interest, but they are the principal sources of nonlinear distortion, overloading, and noise.

a. Vacuum Tubes

The vacuum tube is composed of two or more elements (electrodes) sealed inside an evacuated glass (sometimes ceramic or metallic) envelope. A two-element tube, the diode, can only rectify, changing an a-c signal to d-c, without any amplification. The two elements are the anode (or plate) and the cathode. Some

means of heating the cathode must be included within the envelope, and, in some cases, the cathode is the filament which is heated directly, from an external source of electrical power. In other cases, a separate heating element is used, called a heater.

The introduction of a third element, the grid, placed between the cathode and anode, allows the vacuum tube to amplify by controlling the energy in the secondary circuit by means of the grid voltage. The three-element tube, the triode, is used frequently in both initial and intermediate stages of amplification. Other vacuum tube configurations include the tetrode (two grids), the pentode (three grids), and even a five-grid array (pentagrid). Each of these types has characteristics somewhat different from the basic triode.

b. Transistors

The transistor is a solid-stage device consisting of three segments of semiconducting material, arranged to provide two junctions. The material is either n type, abundant in free electrons, or p type, lacking in free electrons. The transistor is arranged in p-n-p or n-p-n arrays, forming two p-n junctions. Through a solid-state mechanism of charge neutralization, the current through one junction controls the current through the secondary circuit of the device, which is in series with its other junction. Current amplification of a significant magnitude, above 25, can be obtained.

The transistor, like the vacuum tube, may be connected in different configurations, depending on whether voltage or current gain is desired. Both vacuum tube and transistor amplifiers are now being used in physiological monitoring systems, although transistors are becoming increasingly popular and replacing tubes in most applications. Generally, in an airborne monitoring package, the advantages of transistors necessitate their use, whereas in a ground installation, where size and weight are not major considerations, either type of amplifier may be used.

c. Circuit Configurations

Circuit configurations as noted in figure 29 are classified according to the element which is chosen as the common element of the amplifier, i.e., the element which is shared by the input and output circuits. The criteria for selecting certain configurations are much the same for both vacuum tubes and transistors, although the actual values of the various associated parameters may differ. Whenever maximum voltage and/or power gain is needed, the common (grounded) cathode or emitter configuration is generally employed. The common plate or collector (cathode or emitter follower) configuration is employed when high input impedance, low output impedance or high current gain is desired. The grounded grid vacuum tube circuit is employed whenever a high degree of shielding is needed between the input and output circuits, such as in very-high-frequency amplifiers.

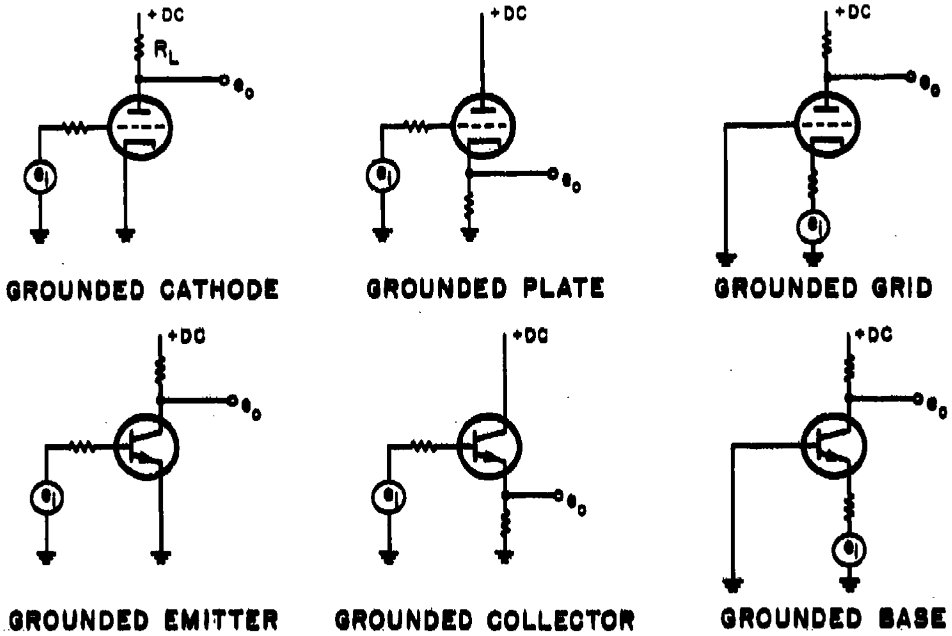


Figure 29. Circuit Configurations for Vacuum Tubes and Transistors

The following are the three main advantages of transistors over tubes:

(1) Small size. The average volume of a small signal transistor is one-fifth or less of the smallest vacuum tube.

(2) Power consumption. About 1 watt of power is used with a vacuum tube just to supply the filament current, which is unneeded in a transistor. This less stringent power supply requirement also decreases the total weight needed for a monitoring system.

(3) Reliability. The transistor has no glass envelope to shatter and no thin filament to break.

The shortcomings of transistors are the following:

(1) Relatively low input impedance. The input impedance of a transistor amplifier may be about 0.1 to 0.01 of that of a vacuum tube amplifier. New semiconductor devices presently being developed and used, notably the field-effect transistor (a device whose control characteristics depend on electric field variations instead of current control characteristics of the normal transistor), are high input impedance devices rivaling the vacuum tube.

(2) Medium gain. A single stage of a transistor amplifier rarely has

a voltage gain of more than 60 compared to an easily obtained 100 from a vacuum tube.

(3) Limited high-frequency response. Certain types of transistors are operable in the hundreds of megacycles while vacuum tubes are obtainable for use in the kilomegacycle band or higher.

(4) Critical temperature dependence. A transistor must be protected against the effects of high temperature, both ambient or internally generated. The latter type can cause "thermal runaway" when voltage bias is used, resulting in destruction of the transistor. Competent circuit design using compensating diodes and thermistors can surmount this difficulty and allow operation at temperatures of 150°F.

The principal transistor shortcoming is a lower voltage gain obtainable from the single active device, generally requiring additional stages of amplification to obtain a specified voltage gain. Internally generated noise in some transistors gives trouble at low signal levels, but this appears to be a manufacturing process control problem. Most of the other limitations are not serious considerations in physiological work.

2. Passive Elements

The passive elements of an amplifier stage are the basic building blocks of electronic circuitry and act to modify and control the operation of the active devices. There are three basic elements, the resistor, the capacitor, and the inductor; all three are shown schematically in figure 30. These passive elements are basically linear devices; i.e., current through the devices is proportional to the voltage across the passive element, without any higher order terms entering into the relationship. Thus the relationship between voltage and current through a resistor, the simplest of the passive elements, is simply

$$\frac{E}{I} = R \quad \text{(Ohm's Law)}$$

This is Ohm's Law, where E is the voltage, I is the current in amperes, and R is the resistance in ohms.

a. Resistance

Resistors are composed of either a carbon or film mixture with carefully controlled resistance and temperature characteristics, or they are made of long pieces of resistance wire wound about a form. The wire-wound resistor generally is used to handle high currents. Resistors range in value from ohms to megohms (millions of ohms).

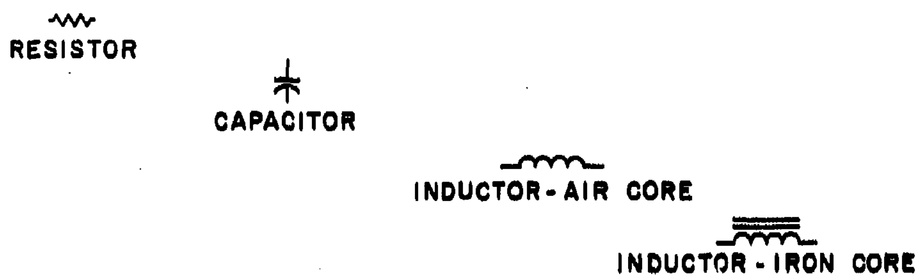


Figure 30. Schematic Representation of Passive Circuit Components

Because of the simple relationship between current and voltage, the pure resistor is not frequency dependent. A value of resistance R is generally the same for d-c voltages or for a-c voltages. However, changes in resistance affect the response of frequency-dependent devices, such as capacitors and inductors.

Resistors often are used in an amplifier to fix the operating points of the active devices (biasing). Rearranging Ohm's Law,

$$E = R\ I$$

it is obvious that if a current flows through a resistor, a voltage drop is produced across it. This characteristic can be used to adjust the various electrode voltages to the desired operating point. In addition, the resistance can reduce the voltage of a power supply to the desired value by means of a voltage drop across the resistor. In addition, the resistance can reduce the voltage of a power supply to the desired value by means of a voltage drop across the resistor.

The resistor also limits current flow, preventing an active element of very low internal impedance from drawing a destructive amount of current from a power source. The output of a voltage amplifier stage is taken across the load resistor, because the tube or transistor controls the current passing through the resistor and thus controls the voltage drop appearing across the resistor.

b. Inductance

An inductor (sometimes called a choke) is an element across which the voltage drop is proportional to the rate of change of the current through the element. Therefore, for a direct-current flow through the inductor, there is no voltage drop (zero impedance) (this is true only for an ideal choke, one that has no internal resistance). As the frequency of the current through the element increases, the impedance (which may be thought of as the resistance to an a-c current) increases linearly. The actual impedance an inductance offers to the current may be computed from the formula:

$$Z_L = 2\pi f L$$

where Z_L is the impedance in ohms, f is the frequency in cps, and L is the inductance in henries. An inductor is basically a coil of wire and the value of the inductance is a function of the geometry, primarily the number of turns and the core material about which the turns are wound. Generally, air core inductors have low inductance values, whereas inductors having ferromagnetic cores may have a high inductance value. Inductors may be either solenoidal or torroidal in shape.

Since the impedance of an inductor increases with increasing frequency, it may be used in conjunction with resistances to form a simple filter. When placed in series with the signal path, it acts as a low-pass filter. When shunted across this path, it acts as a high-pass filter. When used in conjunction with the proper complementary capacitors, rather than resistors, filters may be designed having controlled attenuation characteristics.

Inductors are rated according to their inductance value, current carrying capability, and d-c resistance.

c. Capacitance

A capacitor has a current-voltage relationship which is the inverse of that of the inductor. The current through a capacitor is proportional to the rate of change of voltage across it. As the voltage variation (frequency) increases, the impedance of the capacitor decreases. Expressed mathematically,

$$Z_c = \frac{1}{2\pi f C}$$

where Z = impedance in ohms,
f = cps, and
C = capacity in farads.

The capacitor is constructed of two conductors, often pieces of metal foil, separated by a dielectric material. The dielectric may be paper, ceramic, or mica (used for very high voltage applications). In electrolytic capacitors, thin films of dielectric material are deposited directly upon the conducting foil by a liquid electrolyte.

Capacitors are rated according to their capacity (expressed in microfarads or picofarads), and their maximum voltage. Electrolytic capacitors generally are polarized, with the rated capacitance obtained only if the proper polarity is observed.

Because of its voltage-current relationship, the capacitor acts as a high-pass filter in series with the signal line and as a low-pass filter (high-pass shunt)

when connected across a signal transmission path. Because capacitors do not pass dc, they are used to isolate one stage of amplification from the next, allowing only the signal variation (ac) to be passed to the succeeding stage.

B. Classes of Stages

The main function of a signal modifier is to improve the signal characteristic so that the signal is more useful for display and analysis. The usual task of the amplifier is to obtain voltage, current, or power gain. The inherent characteristics of the active elements of the amplifier stages limit the power gain obtainable for each stage. The total power gain desired is obtained by using successive stages, one amplifying the output of a previous stage. The load for one stage, which may be the following input stage or a display device or transmitter, defines the requirements for one class of amplifier. Input considerations or the character of the signal source may be the factor of importance, dictating another type of amplifier class.

1. Power Stage

The power stage is used when the function of the stage is to supply appreciable power to a load having essentially resistive characteristics. This stage would be used to energize an indicator or oscillograph pen directly, or to modulate the output stage of a telemetry transmitter.

Power stages are classified according to the configuration and number of active elements used. The requirements of the power stage are dictated by the amount of power required, the accuracy of amplification desired, and the type of load being driven by the stage. The single-ended stage is the most basic configuration. A push-pull connection provides more gain and fidelity. The parallel stage, or the cathode or emitter follower stage is used when high power requirements or low-impedance loads are associated with the power stage.

a. Single-Ended Stage

In the single-ended configuration, the input signal is referred to ground and the input stage of the amplifier responds to all signals at the input terminal with respect to the common ground point. The basic amplifier configurations in figure 29 are single-ended connections.

b. Push-Pull or Balanced Stage

For high-power, low-distortion output requirements, the push-pull amplifier configuration (shown in figure 31) is used for the output stage. Two tubes share a common power supply and common output transformer. The inputs to the control grids of each of the two tubes are exactly 180° out of phase, so that the current in one tube is increasing as the other is decreasing, which is the reason for the term push-

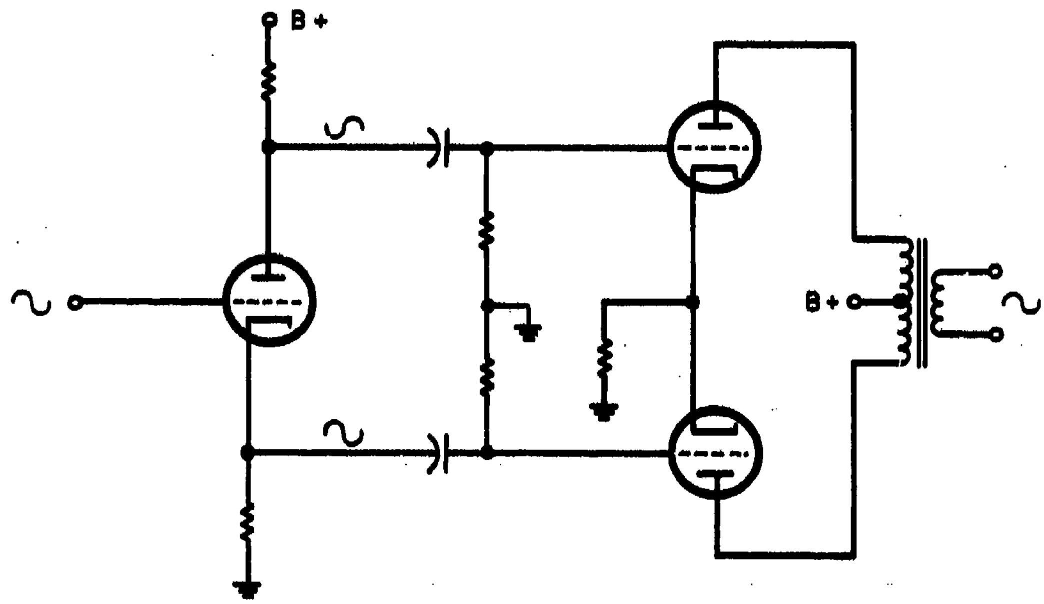


Figure 31. Push-Pull Amplifier Driven by a Phase Splitting Stage

pull. Since the flux produced by the two currents is being added in the transformer core, any even harmonics of the fundamental frequencies that are present in the input signal are cancelled, reducing harmonic distortion greatly.

The out-of-phase input signal to the grids of the push-pull tubes may be supplied by a phase-splitting amplifier stage, as shown in figure 31, or by a transformer with the secondary center tap connected to ground.

The push-pull amplifier is less sensitive to variations in the power supply. A greater percentage of ripple may be tolerated without being reproduced in the stage output.

c. Parallel Stage

If considerable output power is required, paralleling of stages is permissible to effectively double the power. Figure 32 illustrates a parallel stage of two triode vacuum tubes. For still more power, a push-pull parallel combination may be used, both increasing the output power and gaining the advantages of push-pull operation.

d. Current Stage

The current output stage (a special power amplifier) is used when the current through a load having an appreciable reactance must be proportional to input

voltage over a wide frequency range. For example, this stage would be used for driving a magnetic recording head or deflecting a galvanometer indicating instrument. This stage, with a high ratio of driving impedance to load impedance, uses an active element like a pentode, which has an extremely high internal impedance.

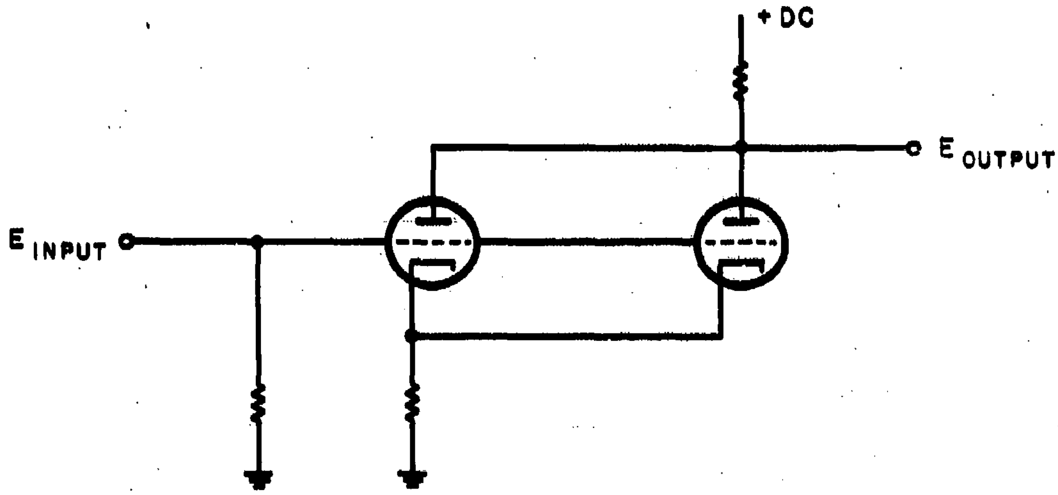


Figure 32. Parallel Output Stage

e. Cathode or Emitter Follower Stage

When another device, such as a meter, recorder amplifier, or telemetry amplifier or oscillator, is the amplifier load, a low-impedance, few-volt-level signal usually is desired to minimize high-frequency losses and loading by cable or following equipment. A grounded plate (cathode follower) or grounded collector (emitter follower) connection of a tube or transistor satisfies these requirements. A linear output is achieved over a very wide frequency band. No voltage gain is present, but there is current gain.

2. Voltage Stage

When the physiological signal is too weak to drive a power stage directly, voltage stages of amplification are used to increase the signal level. In special cases, they are employed as input, front-end amplifiers. Usually, however, they follow an input differential amplifier stage. The first and third pair of tube and transistor configurations in figure 29 are voltage amplifiers.

The input to a secondary or tertiary voltage stage is not as critical in design as the input to a first stage of amplification. When it reaches the higher level stage, the signal has already been amplified and any common mode noise has been rejected (if a differential amplifier was used). For these reasons, the voltage stage may be of single-ended input and output design.

If the previous stage is a carrier amplifier, the voltage stages are tuned amplifiers; the coupling between each stage consists of a resonant circuit tuned for the carrier frequency. In this way, any extraneous noise that manages to be interposed upon the modulated carrier signal is rejected by the tuned coupling network. A tuned, voltage stage amplifier can have very high gain with little noise or instability.

3. Difference or Differential Voltage Stages

In physiological monitoring, the differential amplifier is probably the most important and most used class of amplifier. When the biological signal source is the input to the amplifier, the characteristics of the source must not be disturbed by the act of measuring. In monitoring situations where more than one signal is originating from a single area of the biological subject, interactions between the amplifiers associated with each signal may be encountered. This is especially true if the amplifiers have all of their inputs referenced to a common ground, as is the case with single-ended input amplifiers. Current may flow between amplifier inputs, and these currents may produce voltages which can change or mask the true physiological signals. The differential amplifier input stage allows each signal to be measured independently, because no common ground connection is made between each amplifier input stage, maintaining the integrity of each signal measurement.

The differential amplifier is characterized by two independent input terminals, each referenced to a common terminal, usually amplifier ground. While the inputs are connected to two separate input electrodes, the outputs of the two stages are connected with a common resistor. The output of the differential stage is the difference between the inputs.

The output of the differential amplifier stage also may be a balanced type; i.e., two outputs that are not referenced to ground, but are referenced to each other. In many instances, however, it is desirable to transfer the difference signal from a balanced type of differential stage to a single-ended or grounded stage, so that the output may be amplified further by single-ended voltage stages. The advantage of the differential stage, the common mode rejection characteristic, must be retained in this transfer, which may be difficult to achieve. Figure 33 illustrates two circuits, using tubes and transistors, that are used to accomplish the transfer from a balanced to a single-ended configuration. The latter circuit permits the adjustment of the common mode rejection of the stage to maximize this parameter, and adjusts for variations in active and passive components due to manufacturing variations and changes due to aging.

C. Interstage Coupling and Stage Termination

The methods used to interconnect the individual stages of amplification can determine the frequency response of the amplifier. Although the individual stages have no frequency limitations in the range of physiological signals, the interstage coupling may be a problem.

1. Direct Coupling

Direct coupling of stages (see figure 34) will give frequency response down to zero frequency or dc. The use of direct coupling, however, introduces problems, especially when high gain is necessary in the early stages. Direct coupling usually requires cumbersome and expensive power supplies, particularly with vacuum-tube amplifiers, and a common source of drift and instability is the power supply. The effects of drift and instability are amplified by succeeding stages, causing these extraneous signals to be indistinguishable from the biological signal.

2. Capacitive Coupling

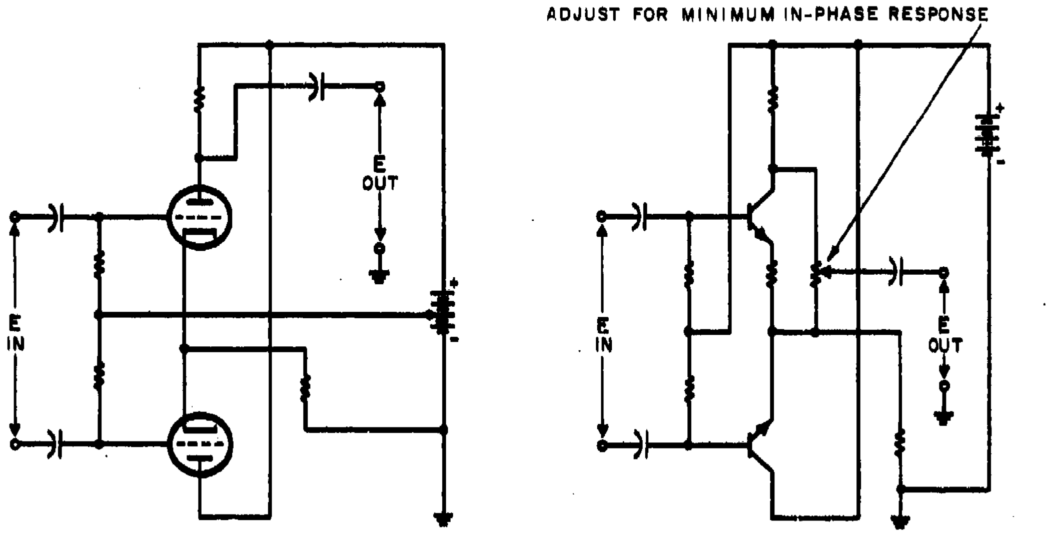
A capacitor is the component used most commonly to couple amplifier stages. The capacitor blocks the flow of dc, eliminating drift effects from the power supply, and permits the use of simple battery and rectifier power supplies. The capacitor is a frequency-dependent impedance component, with the impedance varying inversely with the frequency of the applied signal. Consequently, the capacitively coupled amplifier cannot be used for amplifying d-c or very-low-frequency signals.

Advances in capacitor technology, especially in the fabrication of tantalum capacitors that allow large capacitances in tubular configurations small enough to be used in the miniaturized physiological amplifiers, have lowered the frequency limit to 0.2 cycle per second. At high frequencies, the capacitor impedance approaches that of a short circuit, and the response of the coupling networks presents no limitations in the use of this technique. However, if the size of the coupling capacitor is increased to extend the low-frequency range, trouble may be experienced with blocking effects after voltage pulses (see page). Figure 34 illustrates the basic capacitive-coupling configuration.

3. Inductive or Transformer Coupling

A transformer is used to couple two stages of amplification in special cases where limited frequency bandwidth is desired or where impedance matching is of prime importance. Figure 34 illustrates how the transformer is used to couple stages. The number of windings in the primary and secondary of the transformer is determined by the impedance of the primary and secondary circuits. The ratio of the impedances is proportional to the square of the ratio of number of turns in the primary and secondary (if there are 200 windings in the primary and 400 in the secondary, the ratio is 2:1 and the impedance ratio is 4:1).

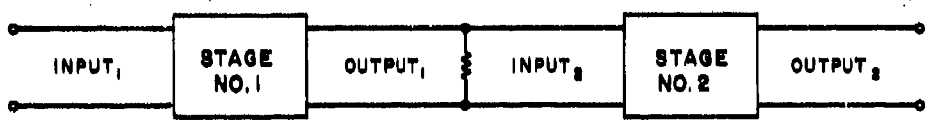
For a transformer to pass very low frequencies (below 10 cycles per second), the core is necessarily large and the construction rather expensive and heavy. For this reason, transformer coupling is seldom used in physiological work, where so many signals of interest are below one cycle per second.



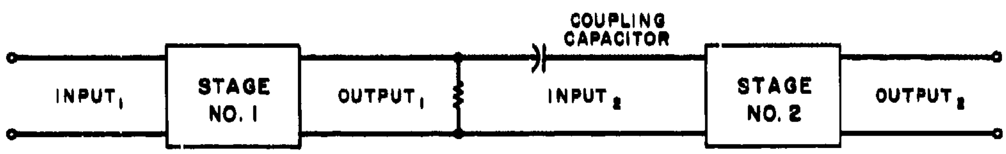
VACUUM TUBE AMPLIFIER
DIFFERENTIAL INPUT - GROUNDED OUTPUT

TRANSISTOR AMPLIFIER
DIFFERENTIAL INPUT - GROUNDED OUTPUT
(HIGH COMMON MODE REJECTION)

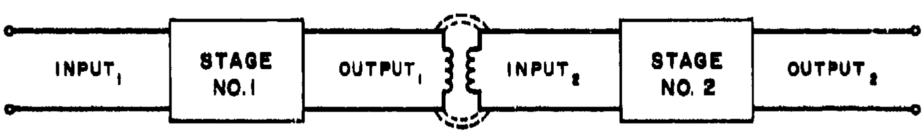
Figure 33. Differential Amplifier



DIRECT COUPLING



CAPACITOR COUPLING



TRANSFORMER COUPLING

Figure 34. Interstage Coupling

At r-f frequencies, transformer coupling is used extensively because size and weight are not a problem, since air core transformers instead of iron core transformers may be used; it is possible to construct a tuned circuit most responsive to the band of frequencies under consideration; and transformer coupling provides a convenient method of splitting a signal into in-phase and reverse-phase components for push-pull input and output stages.

D. Special Systems

Special coupling techniques are used when capacitive coupling does not provide a low-frequency response of sufficient range in a monitoring system amplifier and drift considerations preclude direct coupling. The most common approach involves the use of a modulated carrier. In this technique the actual d-c signal from the input terminals is used to modulate or control the amplitude of an a-c signal that the amplifier stages can amplify with complete fidelity. The original signal is then recovered by demodulation.

1. Chopper Amplifiers

The chopper amplifier converts the low-frequency or d-c input signal into an a-c signal (chopping effect), amplifies the a-c signal to the desired level, and converts the a-c signal to its original low-frequency or d-c form. The d-c signal resulting from the conversion must be proportional in magnitude to the d-c or low-frequency signal being measured (see figure 35).

Electromechanical choppers convert the dc to ac at the input stage. Basically, electromechanical choppers are miniature relays designed to open and close a set of contacts with each alternation of an applied a-c supply voltage. The chopper can be connected between the input grid of the a-c amplifier and ground, and each time the chopper contacts close, the input signal is reduced from its d-c value to zero. The result is a pulsating d-c signal with an amplitude proportional to the applied d-c signal. By using resistance-capacitance (RC) coupling in the amplifying stages, the pulsating d-c signal appears in the output as an amplified a-c signal. The amplification is simple and stable.

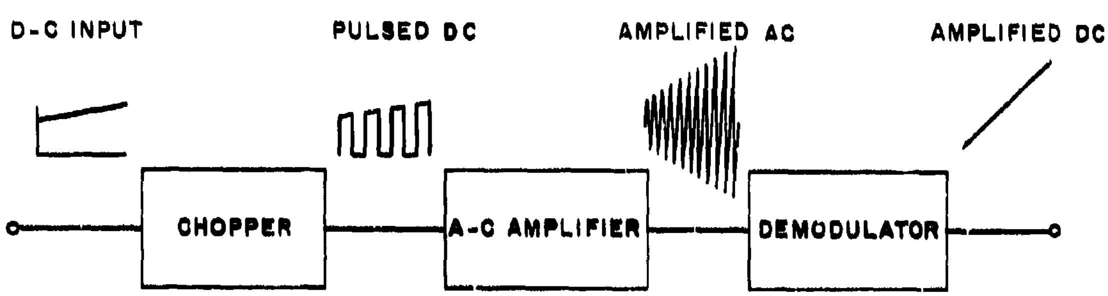


Figure 35. Chopper Amplifier

After amplification to the desired level in the intermediate stages, the a-c signal is fed to the demodulator where it is rectified and smoothed. It then appears at the output stage as an amplified version of the original low-frequency or d-c input signal.

Available chopper amplifiers are capable of supplying drift-free operation with input signals as low as 1 millivolt. The amount of distortion caused by nonlinear circuit operation in these amplifiers is less than 1 percent.

Electromechanical choppers are unsuitable for use in extreme environmental conditions, especially those of acceleration and shock. Typical troubles are contact chattering, armature hold from direct acceleration effects, and mechanical warpage and breakage. A solid-state chopper, which essentially is an electronic switch, provides the same function as the electromechanical device: switching the amplifier input between the measured d-c level and ground. Their stability under acceleration and shock makes them usable for airborne and centrifuge chamber instrumentation. However, problems may be encountered under variable temperatures.

2. Carrier Amplifiers

The principle of operation of a regular carrier amplifier is similar to that of a chopper-type amplifier. Carrier amplifiers do not amplify the physiological signal at its original frequency, but instead amplify a higher frequency carrier signal whose amplitude has been varied in proportion to the input signal. A block diagram of a basic carrier amplifier is shown in figure 36. The low-frequency physiological input signal modulates the amplitude of a carrier signal produced by the carrier frequency oscillator at the input stage. The input to the a-c amplifier, therefore, is a signal at the frequency of the oscillator, but with an amplitude which varies directly with the amplitude of the low-frequency input signal. This type of signal is referred to as a modulated carrier. After the modulated carrier is amplified in the input and intermediate stages, it is fed, along with a reference signal from the carrier frequency oscillator, to a phase-sensitive demodulator. The output of the demodulator is an amplified reproduction of the low-frequency input. Carrier amplifiers are characterized by extremely low drift, and freedom from transient blocking effects.

3. Bridge Amplifiers

When the transducer in a physiological monitoring system is a variable resistance, inductance, or capacitance device, a bridge circuit usually is used because it increases sensitivity, provides temperature compensation, and minimizes the effects of signal leads. In a strain gage application, the strain gage is connected as one leg of a bridge. The bridge amplifier chassis usually contains a power supply (for bridge excitation), three legs of the bridge, and an a-c or d-c amplifier to boost the small variation in the bridge output to a usable level.

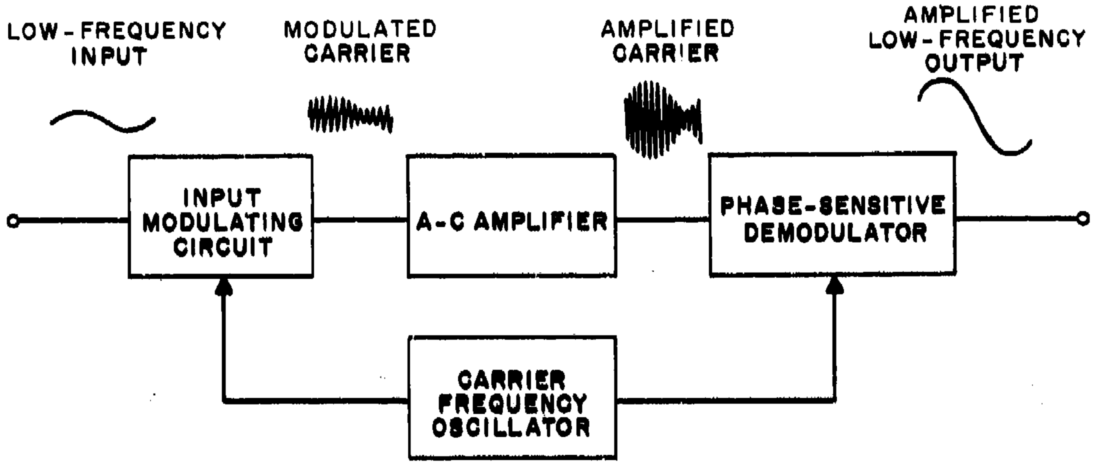


Figure 36. Carrier Amplifier

Figure 37 illustrates a Wheatstone Bridge configuration used for resistance determinations. For a variable inductance or capacitance transducer, a reactive bridge is used. Figure 37 shows a Wien Bridge used for a variable capacitance transducer system and a Maxwell Bridge used for variable inductance transducers. For both these bridges, a-c excitation must be used.

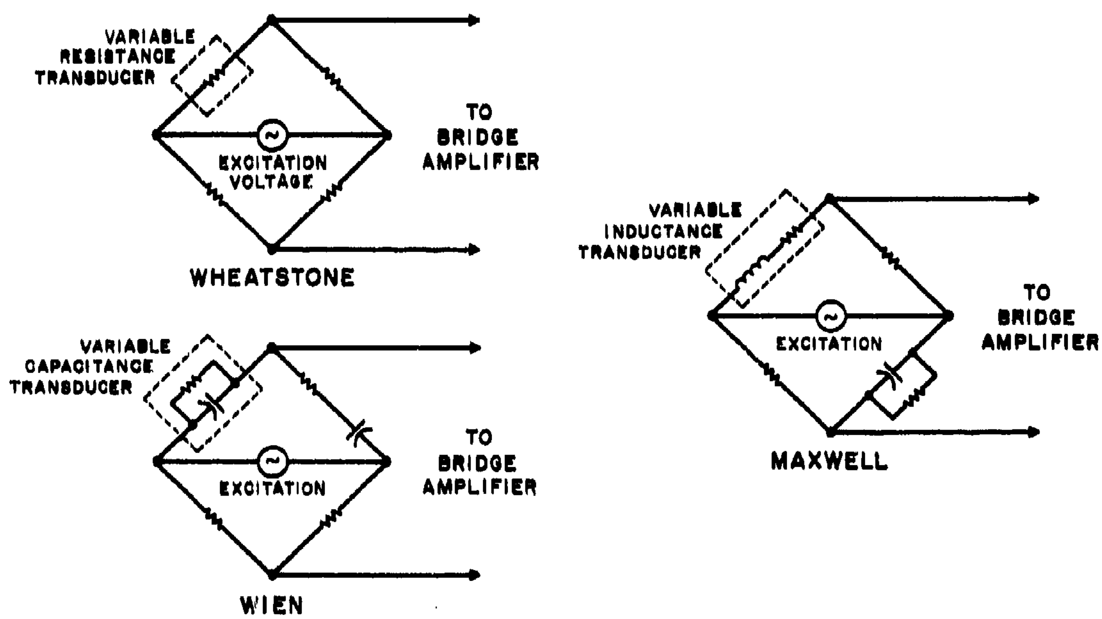


Figure 37. Bridge Configurations

When a-c excitation is used for a bridge, the amplifier may be considered to be a carrier amplifier, with the carrier frequency the excitation frequency. Thus the d-c response of the amplifier is not important. The most important criterion when choosing the excitation frequency is that the carrier (excitation) frequency be much higher than the highest frequency present in the phenomena being measured by the resistance, capacitance, or inductance transducer. Ten times the highest harmonic frequency of the modulating signal usually is sufficient to ensure that, when the carrier frequency is removed from the signal in the demodulator-filter circuits, no information from the modulating signal is lost.

The output of the amplifier may be in the form of an amplified a-c, d-c, or modulated carrier waveform. The demodulation circuit often is included within the amplifier package, however, producing an integrated circuit requiring an input transducer and power source, and supplying the d-c output as a replica of the input.

For versatility, some bridge amplifiers have provisions for disconnecting the internal bridge arms, providing an a-c amplifier without a bridge input, which can be used with a differential transformer transducer for monitoring purposes.

4. Ionization Tube Signal Modifier

A unique signal modifier that can be considered a combination carrier amplifier and demodulator system is shown in figure 38. Although this unit cannot be

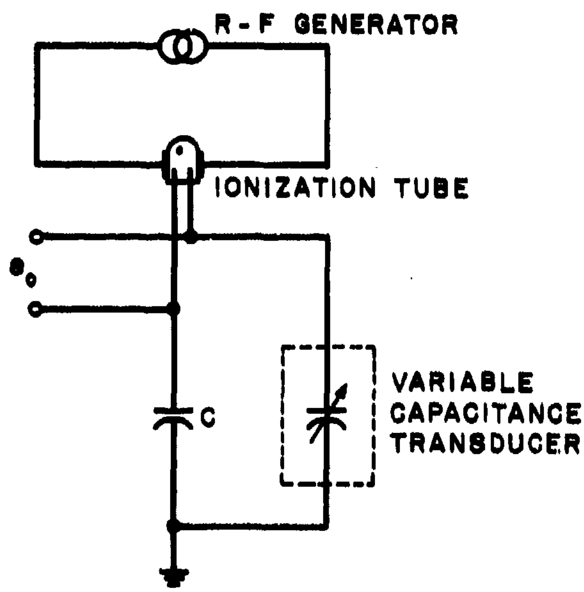


Figure 38. Ionization Tube Signal Modifier

classified as an amplifier (no voltage or current is amplified), functionally this appears the logical place to discuss it, since it is the initial signal modifier following the transducer.

The heart of this signal modifier is the ionization tube (ref. 23), which is a small, partially evacuated glass vessel containing an inert gas (usually neon) at about 15 mm of mercury pressure. The tube has two plates positioned external to the glass envelope and two electrodes inside the envelope are in direct contact with the inert gas. When an r-f generator is connected to the external plates, the gas is ionized by the resultant field. This signal modifier is used with a variable capacitance transducer. The variable capacitor, connected across the electrodes, controls the potential across the electrodes resulting from the electric field, and the output voltage is proportional to the transducer variation.

The d-c voltage variation at the electrode terminals is as high as several volts per picofarad of variation of the transducer, and does not require stages of voltage amplification for eventual analysis or recording.

OTHER SIGNAL MODIFIERS

The information contained in the output signal of a physiological transducer may not be in a form that is adaptable to analysis, and therefore must be processed by modifiers before significant data are available to the experimenter. Modifiers also are used when more than one category of information is obtained from the same basic signal. A singular example of this situation is the output signal obtained from a chest expansion transducer. By integrating this signal, a measure of inspired air volume is obtainable; by a series of more complex operations on the signal, a respiration rate is obtained as noted in figure 45. The more commonly used modifiers are described below.

1. Attenuators

An attenuator is an arrangement of components (usually passive), that performs a function opposite to that of an amplifier. An attenuator decreases the signal amplitude or power. It is used primarily in physiological measuring systems to compensate for variations in biological signal levels obtained from different subjects or different conditions. Amplifier stages are constructed with components having fixed values. Therefore, amplification factors are fixed, and only by deleting an entire stage of amplification can the gain be decreased. To avoid this cumbersome method of adjusting gain, attenuators of fixed or variable values are inserted either at the input stage or between stages, controlling the overall gain of the amplifier.

In physiological work, attenuators generally are used only in voltage stages and in comparatively simple circuit arrangements, using resistors as the passive components. Attenuators also may be used as balancing elements in balanced or differential ampli-

fiers. In very-high-frequency ranges (greater than one megacycle), attenuators may be composed of critical arrangements of capacitive or inductive components. Using transmission lines to interconnect elements of a monitoring system requires special impedance matching techniques, and attenuators are used in this application.

Most simple attenuators are voltage dividers (see figure 39). The input voltage to the device is applied to the two resistors in series, while the output is taken across the lower resistor. Depending upon the relative magnitudes of the two resistors, the voltage divider attenuator may provide any percentage of the input voltage. By using a variable resistor, as shown in figure 39, variable attenuation or gain control is possible.

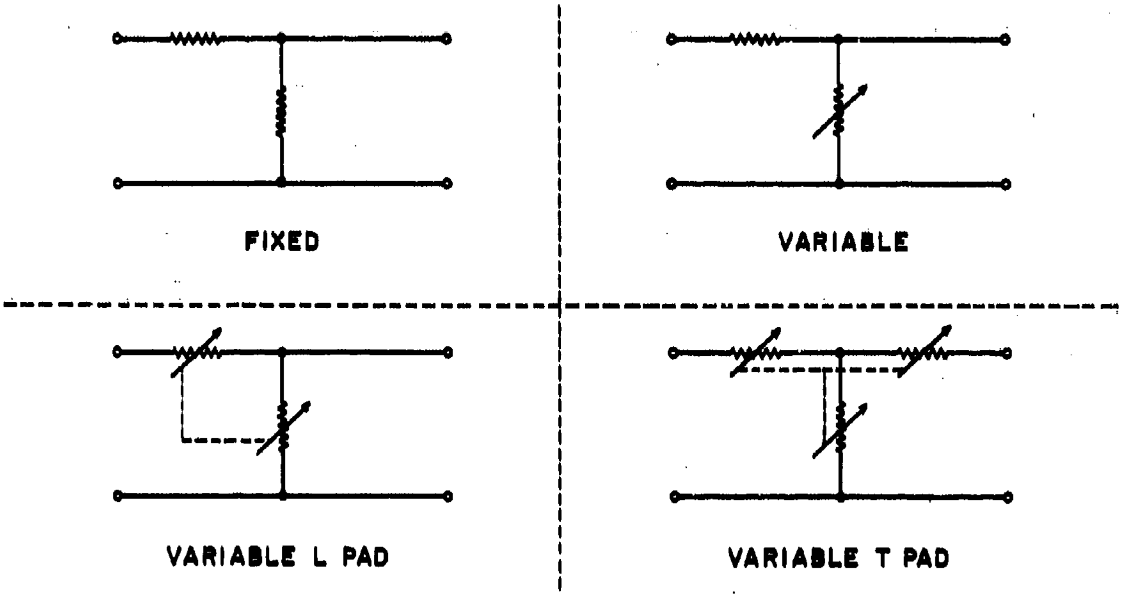


Figure 39. Attenuators

Two slightly more complex attenuators, shown in figure 39, are used when constant input and output impedance considerations are important. The variable L pad and T pad, so called because of the shape of the resistor configuration, depend on ganged or linked resistors whose values all change simultaneously upon rotation of a single shaft. The L pad provides a constant impedance to the input circuit, but allows variation in the output circuit impedance. The T pad, however, maintains constant input and output circuit impedances for all settings of the attenuator control.

II. Modulators

Modulation is the introduction of information into a reference signal by changing some parameter of the reference signal. Two types of reference signals are used: one is called ~~a~~ continuous carrier signal and the other a pulse train. The continuous carrier is used more often, but the pulse train is making inroads into this preference.

The types of modulation used with a carrier signal are amplitude modulation, frequency modulation, and phase modulation. Amplitude modulation is the easiest to produce and detect, but it has some severe limitations. (Figure 40 shows the amplitude modulation of both d-c and a-c input signals.) The amplitude of the carrier varies in proportion to the amplitude of the modulating signal; the rate of change is proportional to the frequency of the modulating signal. The carrier frequency must be appreciably higher than the highest frequency contained in the modulating signal. One basic limitation in the amplitude-modulated system is that any extraneous noise produced by nearby electrical disturbances is superimposed upon the desired signal and detected with the modulating signal. Amplitude modulation techniques that conserve the average power needed to produce the modulated signal include suppressed carrier, single sideband, and vestigial sideband modulation.

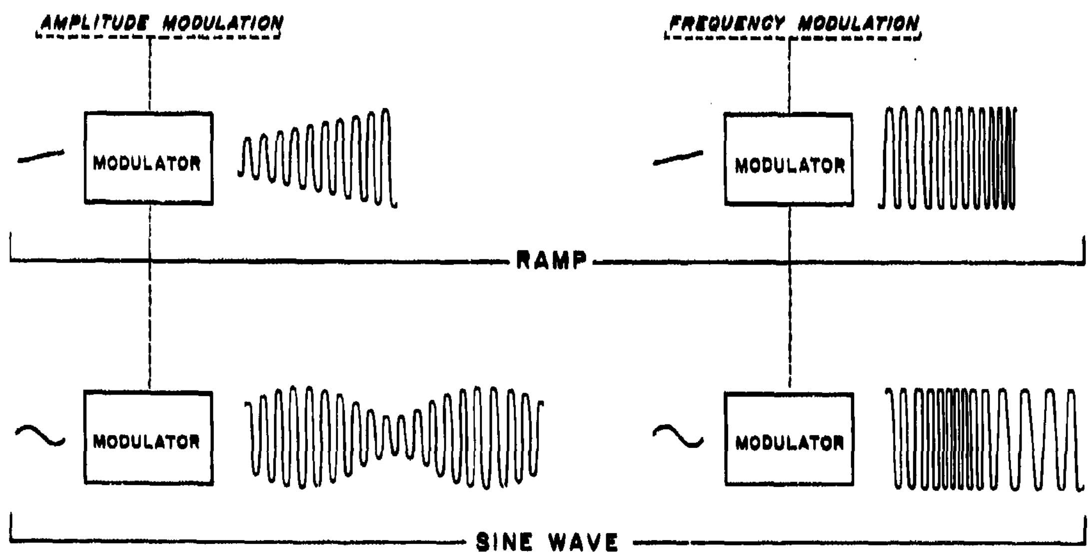


Figure 40. Amplitude and Frequency Modulation Waveforms

In frequency-modulated systems, the frequency of the carrier signal deviates from its nominal center frequency by an amount directly proportional to the amplitude of the modulating signal. The rate at which the deviation occurs is dependent on the frequency components of the modulating signal. (Figure 40 shows the frequency modulation of a carrier signal.) In this system any variation in the amplitude of the carrier has no significance, since all information is contained entirely in the frequency change. Frequency modulation techniques hinge about the variation of a reactance magnitude with the modulating signal. The variable reactance then becomes the frequency-determining element of the oscillator generating the carrier frequency. Consequently, the carrier frequency varies with the modulator signal amplitude.

In phase-modulation systems, the relative phase of the carrier signal varies in proportion to the amplitude of the modulating signal. This type of signal is detected with essentially the same types of circuits that are used in frequency modulation. Phase modulation also depends on the variation of a reactance with the modulating signal. However, the reactance is part of a phase-shifting network that varies the relative phase of the carrier frequency output.

There are three types of modulation used with a pulse train: pulse-amplitude modulation, pulse-position modulation, and pulse-duration modulation. Illustrations of these pulse-modulation techniques are shown in figure 41. Pulse-amplitude modulation (PAM) varies the height of a pulse in proportion to the amplitude of the modulating signal. Pulse-position modulation (PPM) varies the time position of a pulse relative to a reference pulse in a pulse train, the duration of the intervening time period being proportional to the amplitude of the modulating signal. Pulse-duration modulation (PDM) also called pulse-width modulation (PWM), refers to the varying of the duration of the pulses in proportion to the modulating signal.

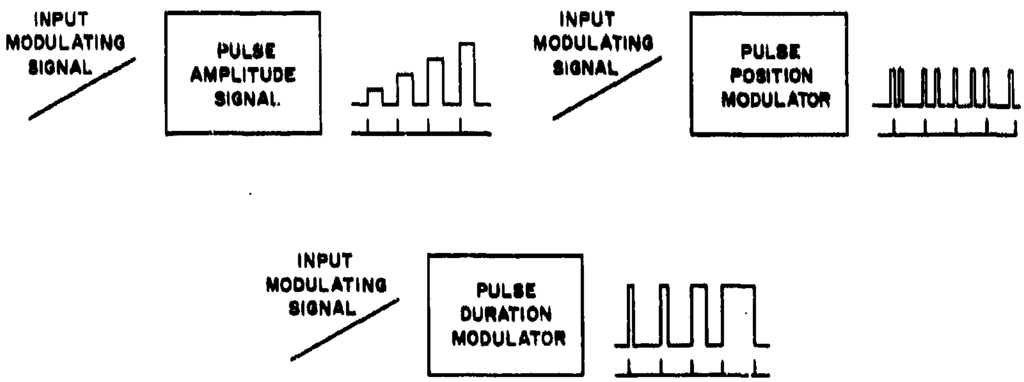


Figure 41. Pulse Modulation Waveforms

III. Demodulators

To extract information contained in a modulated signal waveform, a type of signal modifier called a demodulator is used. This device is necessary because the signal that is transmitted by either wire or radio often contains not only an intelligence signal, but also signal components that are not related to the desired information. A carrier type of transmission is common, and the carrier signal component (a high-frequency signal) is not used in data analysis and recording. The demodulator removes the unwanted carrier signal and extracts the information.

The degree of complexity of demodulation circuitry depends on the type of modulation used. An amplitude-modulated signal is demodulated rather simply. The carrier frequency is removed and the signal intelligence is made available by a simple

rectifier circuit, as shown in figure 42. The input is the modulated carrier frequency and the output is a d-c signal identical with the modulating waveform.

If frequency modulation is used in the monitoring system, the demodulator must develop a d-c voltage proportional to the frequency shift of the carrier frequency in order to duplicate the original modulating signal. Two circuits are in general use for this function: the ratio detector and the phase discriminator. Figure 42 illustrates these two circuits.

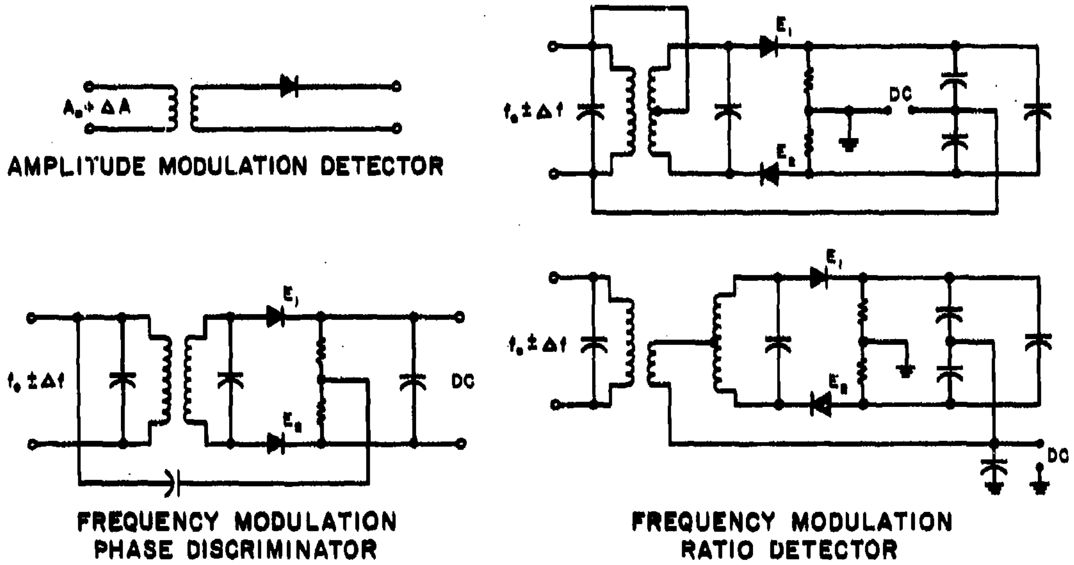
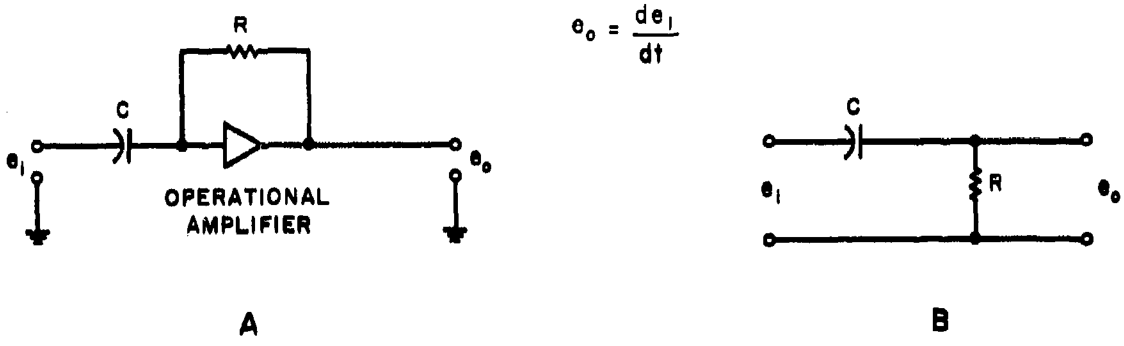


Figure 42. Demodulator Circuits

The ratio detector is relatively insensitive to variations in signal amplitude, as contrasted to the phase discriminator whose output is directly affected by signal magnitude. The phase discriminator therefore must be preceded by a limiting stage for proper operation in a frequency-modulated detection system.

In the ratio detector, the output is derived from the bridge arrangement formed by the two capacitors and two resistors. The sum of the two voltages is held relatively constant by the large capacitor connected across the rectified output, providing a large degree of immunity to rapid amplitude variations.

The phase discriminator diodes are connected to provide a difference voltage output proportional to changes in signal frequency. For an increase in frequency, a positive d-c voltage is obtained, while for a decrease in frequency, a negative signal is produced. If no frequency shift is present, the difference is zero (as desired).



$$e_o = \frac{de_i}{dt}$$

Figure 43. Differentiator Circuits

IV. Differentiators

Differentiation is the deriving of the rate of change of a quantity with respect to a change in another quantity on which it is dependent. An example of this is the rate of change of a voltage in a circuit with respect to time. A circuit that produces an almost ideal differentiated output is shown in figure 43(A). It consists of an input capacitor, an operational amplifier, and a feedback resistor. An operational amplifier is one that is directly coupled, has very high gain (about 1,000,000), has a very high input impedance and inverts the input 180°. By connecting the capacitor and resistor as shown, the output will be the derivative (the differentiated signal). Figure 44 shows the outputs of this ideal differentiating circuit for various inputs.

A simpler circuit for obtaining the derivative of an electrical signal, although it does not produce as ideal a response, is shown in figure 43(B). This circuit produces the derivative because the current through a capacitor is proportional to the derivative of the voltage across it. Since the resistor has the same current as the capacitor passing through it and the voltage across the resistor is directly proportional to the current

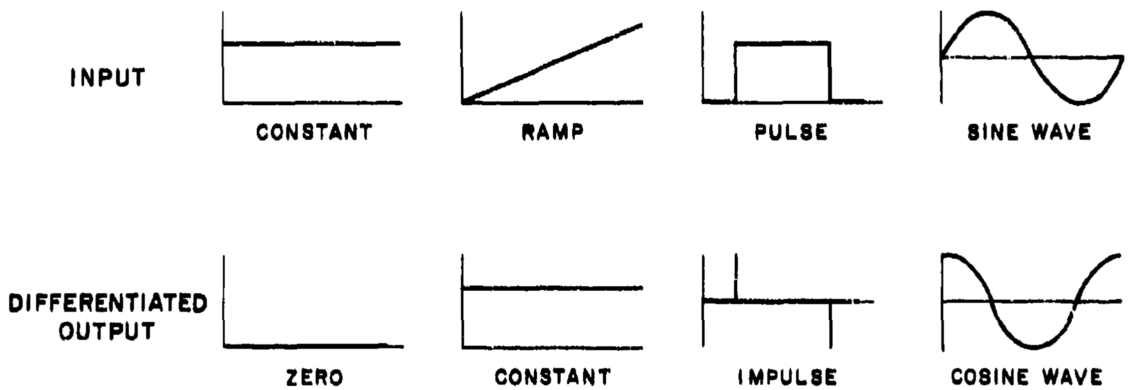


Figure 44. Differentiated Waveforms

through it, the voltage across the resistor is the differentiated (approximately) response of the input voltage.

Differentiated signals are used mainly when the rate of change of a physiological signal is of interest, or when it is desired to eliminate slow drifts. Figure 45 indicates another application in which input signals received from a chest expansion transducer are converted into a series of pulses whose number per unit time is the rate of respiration.

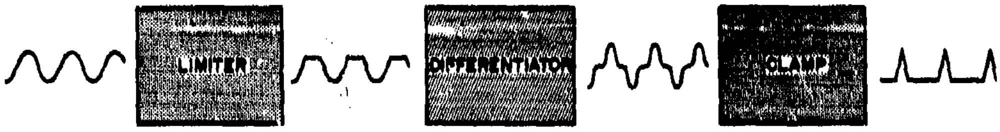


Figure 45. Respiration Rate System Block Diagram

V. Integrators

An integrator is a circuit whose output is the algebraic sum of input function variations with respect to time. Figure 46 shows a few examples of the ideal integration of various input signals. An operational amplifier, with external components connected as shown in figure 47(A), is a circuit that provides almost ideal integration. A simpler though not as accurate method of integrating a signal is illustrated in figure 47(B). The voltage across a capacitor is proportional to the integral of the current flowing to it.

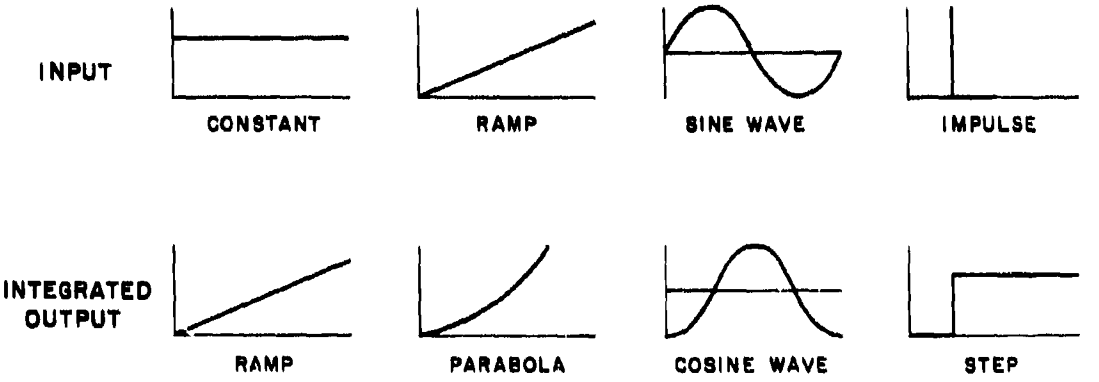


Figure 46. Integrated Waveforms

Integrating circuits are highly useful for converting a signal composed of pulses representing the occurrence of phenomena (e.g., respiration excursions) into a d-c output signal whose magnitude is proportional to the frequency of the input pulses. Therefore, in the circuit of figure 48, after a time sufficient for the capacitor to charge

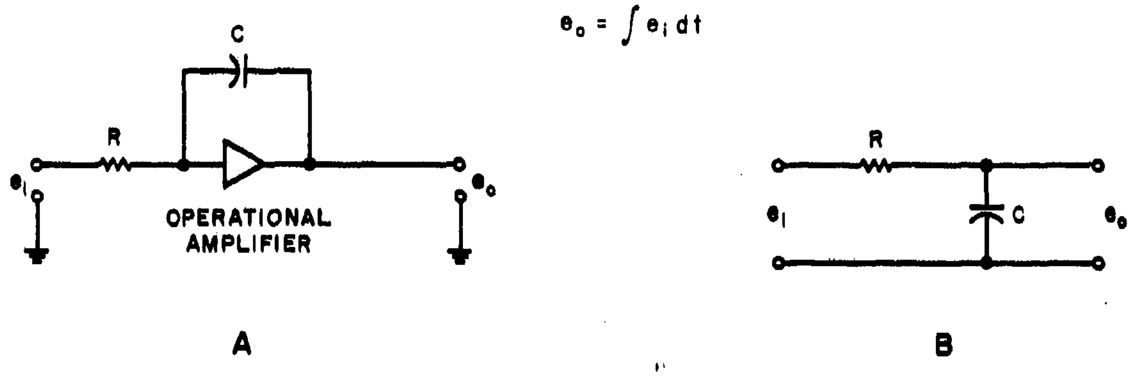


Figure 47. Integrator Circuits

to its initial level, the output level amplitude may be observed on a meter that is calibrated directly in respirations per minute.

VI. Filter Networks

Filters are combinations of passive elements (resistance inductance capacitors) arranged so as to provide desired frequency characteristics in circuits, to remove unwanted frequency components, to isolate portions of circuits from others, and to create self-generating networks (oscillators). Filters usually are classified as low pass, high pass, band pass, and band stop. Figure 49 shows examples of four types in constant K and series m-derived configurations. (Constant K and series m-derived refer to analytic design techniques used to obtain filter networks of desired frequency characteristics.) The filters derived using the constant K method do not provide as sharp a cutoff at the critical frequencies, but the attenuation continues to increase with frequency. The m-derived filter has sharper cutoff characteristics, but the attenuation then decreases somewhat.

Besides introducing attenuation, filters also may produce phase shift. Sometimes they are used to correct unwanted phase shift caused in transmission and amplification, and also in oscillator design. Figure 50 shows a resistive-capacitive phase-shift oscillator. The three filter sections each shift the phase of the signal by 60°. The output of the filter section therefore is 180° out of phase with respect to the output of the

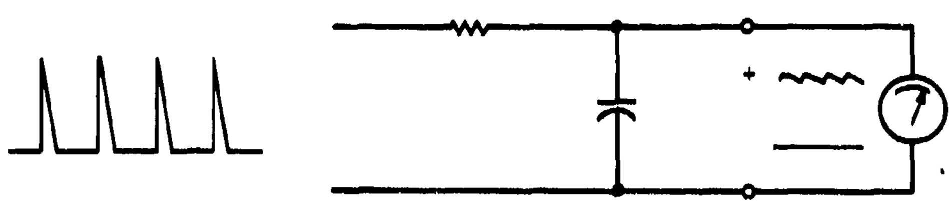


Figure 48. Respiration Rate Indicator Circuit

amplifier stage. When this shifted output is connected to the input to the amplifier stage, the circuit becomes regenerative and oscillation follows.

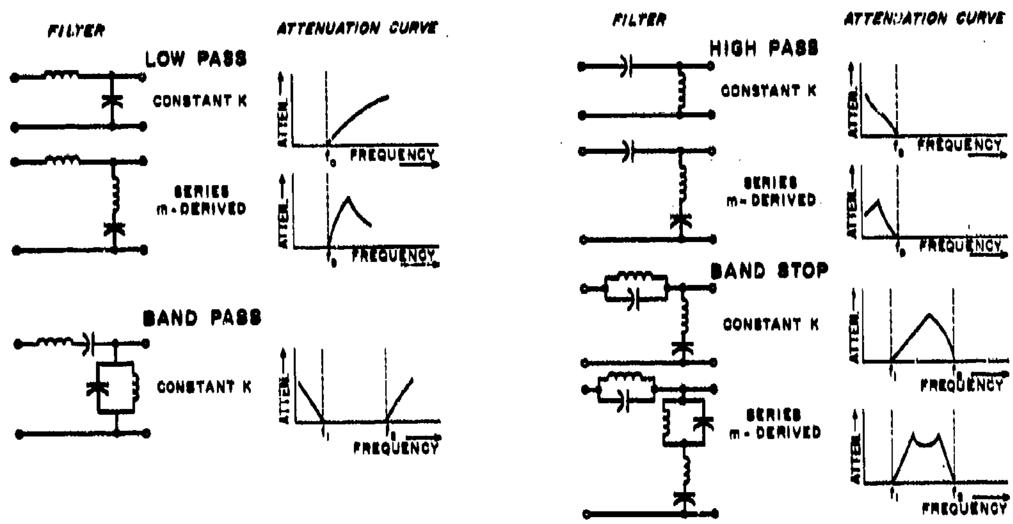


Figure 49. Filter Networks

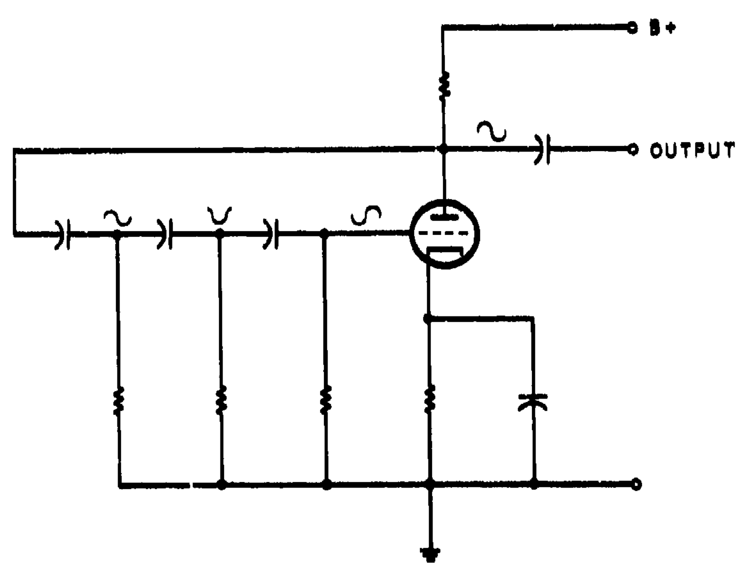


Figure 50. Resistance-Capacitance Oscillator

VII. Limiters

A limiting circuit prevents the amplitude of the output signal from increasing beyond a preset amplitude. This type of circuit, therefore, essentially introduces a non-linear response in a system: below the limiting amplitude, the circuit output may be linear with respect to the input, but voltage peaks above the cutoff point are removed.

A limiting circuit is extremely useful in frequency-modulated systems. Since the information in frequency-modulated systems is contained in the frequency shifts of the signal, the relative amplitude of the signal is unimportant, as long as it is great enough to drive the detector portion of the system adequately. One or more stages of amplitude limiting precede the detector or demodulator stage, especially if the detector is a phase discriminator. Figure 51 shows a typical limiter stage in a frequency-modulated system. The circuit operates on the principle of an oversaturated pentode amplifier. The biasing on the tube is so arranged that the lowest expected signal amplitude output from the intermediate-frequency stage drives the tube into saturation. The resistive-capacitive network in the grid input of the limiter stage tends to maintain the voltage to the grid at a constant value, so fluctuations and sharp spikes in the input signal input do not affect the output of the limiter.

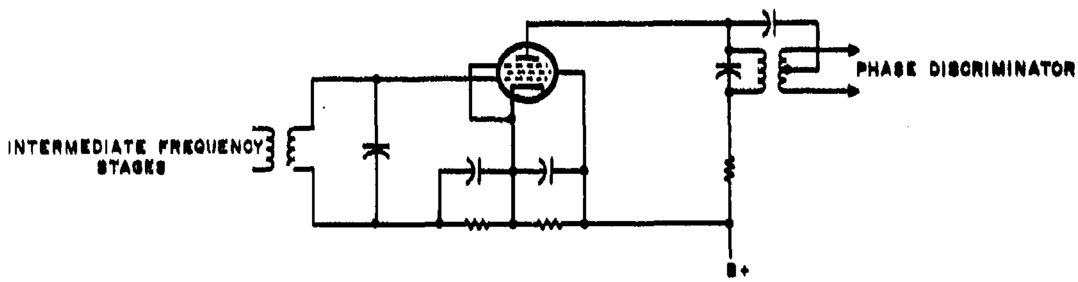


Figure 51. Limiter Stage, Frequency Modulation System

Simple diode limiting also may be used, since a diode does not conduct if the cathode is positive with respect to the anode; therefore, positive peak clipping, negative peak clipping, or positive and negative peak clipping are possible by biasing the diodes and connecting them in different configurations, as shown in figure 52. If Zener diodes are used in this application, as shown in the lower right limiter in figure 52, bias voltages are unnecessary.

VIII. Pulse Generators

It often is desired to produce a pulsed voltage waveform for timing references, amplitude calibration applications, and coincidence gating techniques. Pulses are generated repetitively, as is the case for a multivibrator or oscillator, or they are generated singly, as they would be from trigger signals.

A pulse generator that produces a constant pulse train output is usually some form of a relaxation oscillator. A relaxation oscillator can be as simple as a glow discharge tube circuit or a free-running multivibrator, as shown in figure 53. The circuit selected would depend on the type of output desired and the type of components available. In all cases, however, the basic timing of the waveform generated is determined by a resistor-capacitor combination. If the output pulse is not precisely what is desired, appropriate differentiating and integrating circuits can shape the waveform to the desired characteristic.

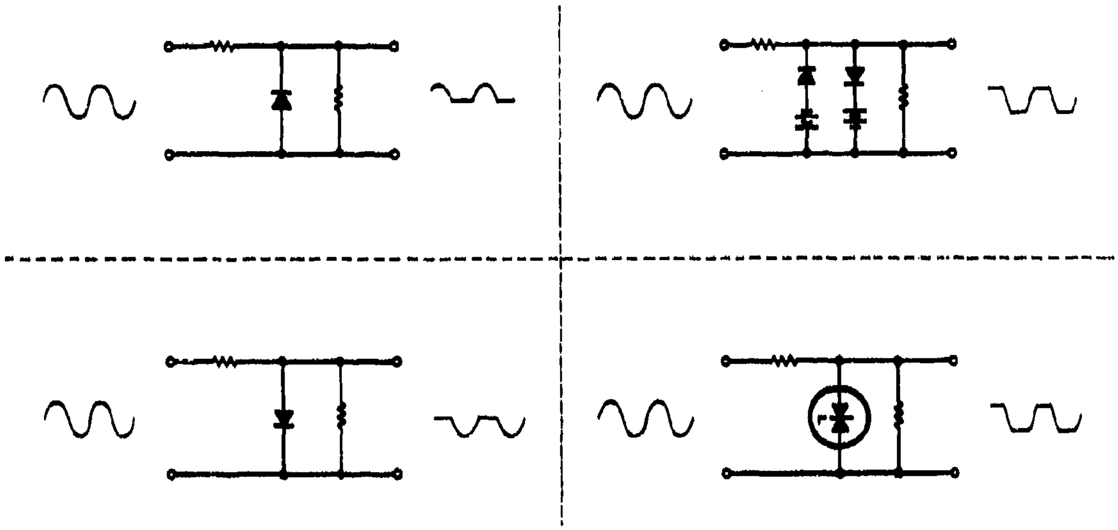


Figure 52. Diode Limiters

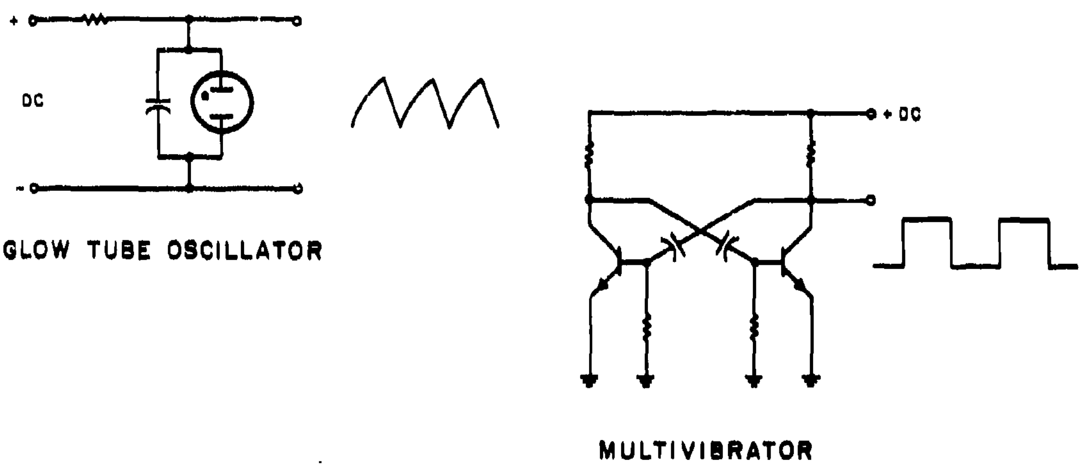


Figure 53. Relaxation Oscillators

A simple case of the generation of single pulses was discussed under differentiation, where the trigger for the pulse was the respiration signal and the pulse was formed from the respiration signal itself.

IX. Triggering Circuits

Triggering circuits produce output signals upon the receipt of an input trigger signal. The output signal may start immediately upon the arrival of the input or it may be delayed a specified time. The output waveform is determined by the type of trigger circuit, and it may have a constant amplitude, constant pulse duration, or other specified characteristics. For example, the Schmitt trigger produces a constant amplitude pulse for the period that the input signal exceeds a specified value; therefore, the circuit can be used in counting circuits, integrating circuits, and control circuitry.

Section III

PRESENTATION DEVICES

GENERAL

This section describes the components used in a physiological monitoring system for the display and recording of physiological data. These devices -- meters, oscilloscopes, and graphic recorders -- convert analog voltages obtained from other components in the monitoring system into a scalar indication, a decimal digit display, or a graphic trace of the voltage in a form that can be understood by the operator.

In many applications, such as the monitoring of body temperature or respiration, the voltages accepted by the presentation device are slow-varying d-c signals, and, if a recording is not required, a simple panel meter is adequate for an instant-readout of the monitored function. If several parameters are being monitored simultaneously or if inputs vary at more than 1 or 2 cycles per second, some form of recording apparatus is necessary. Also, when the pattern of the variations measured is significant, some form of waveform display and recording is essential.

INDICATORS

The indicators of interest in physiological monitoring are essentially voltage amplitude or level indicators. There are two basic types: the panel meter (indicating dial) and the digital indicator. (The oscilloscope is primarily a high-frequency, waveform display device and is treated separately following the discussion of instruments for graphic recording.)

1. Panel Meter or Indicating Dial

The basic movement of a panel meter is that of the D'Arsonval galvanometer.* The galvanometer consists of a small current-conducting coil that is free to move within the field of a permanent magnet. (See figure 54.) The voltage to be measured is applied to the coil, causing current to flow in it. An electromagnetic field results in the coil, which reacts to the permanent magnetic field in such a way as to turn the

---

*The D'Arsonval galvanometer is the most commonly employed meter movement in the range of signals of interest in physiological monitoring. Less common is the electrostatic-type meter movement. There also are meters built around moving-magnet systems, but these are normally designed for high-current measurements (above 50 milliamperes) and industrial applications.

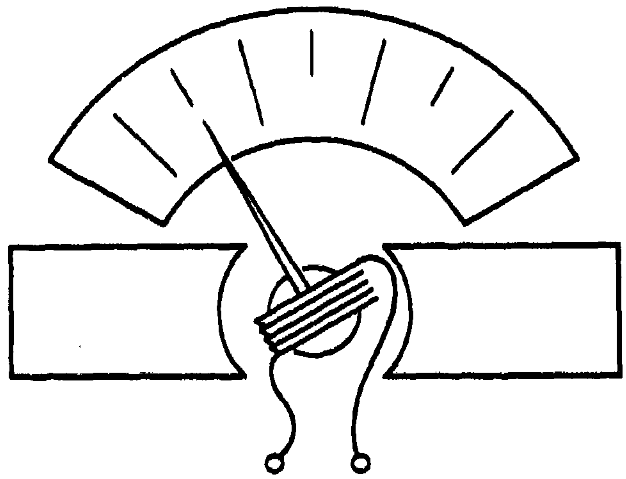


Figure 54. Basic Galvanometer Movement

coil. A dial or pointer attached to the coil travels across a scale, and the deflection of the pointer is proportional to the voltage or current being measured.

The coil is wound around a small magnetic core, and the core is mounted in the permanent-magnet field on jeweled or hard-steel bearings. To keep the mass of this moving system low, the number of coil turns on the core is minimized; this limits the full-scale response of typical meter movements to very low currents, from 0.003 to no more than 50 milliamperes.

To measure higher ranges of voltage or current additional resistors are connected to the coil of the meter.* A multirange instrument is possible if resistors of several values are connected to the coil in various combinations by a switch. Panel meters, however, usually are intended for measuring voltages or currents where the approximate value is known, and therefore are designed for a single range of measurement.

The response of a properly constructed moving-coil meter is linear. Nonlinear response (e.g., logarithmic) is possible by altering the geometry of the permanent magnets used in the meter, or by using nonlinear resistive elements or rectifiers as shunts in the coil circuit. In the latter instance, increasing current or voltage produces increasing current flow through a shunt, rather than through the coil. Such a response is

*To measure current, the meter is placed in series with the circuit, and, to increase the measurement range, shunting resistors are connected across the meter. For voltage measurement, the meter is shunted across the potential sources; in this case, multiplier resistors are connected in series with the meter.

useful for extremely wide-range response in a single-range instrument. The shunts serve the additional function of overload protection; as the measured signal level approaches full-scale deflection, current rises or transient surges beyond that level would be shunted around the coil, rather than overdriving it and causing a possible burnout.

The mechanical characteristics (inertia, etc) of the moving-coil meter limits its frequency response. The usual meter cannot faithfully follow an electrical signal that changes at a rate greater than a few cycles per second. The response time is not particularly important for panel indicators since the observer cannot follow pointer deflections at even this low rate. However, when galvanometer movements are used to operate direct-writing recorders, this limited frequency response must be considered. (Also see the discussion of optical galvanometers on page 105.)

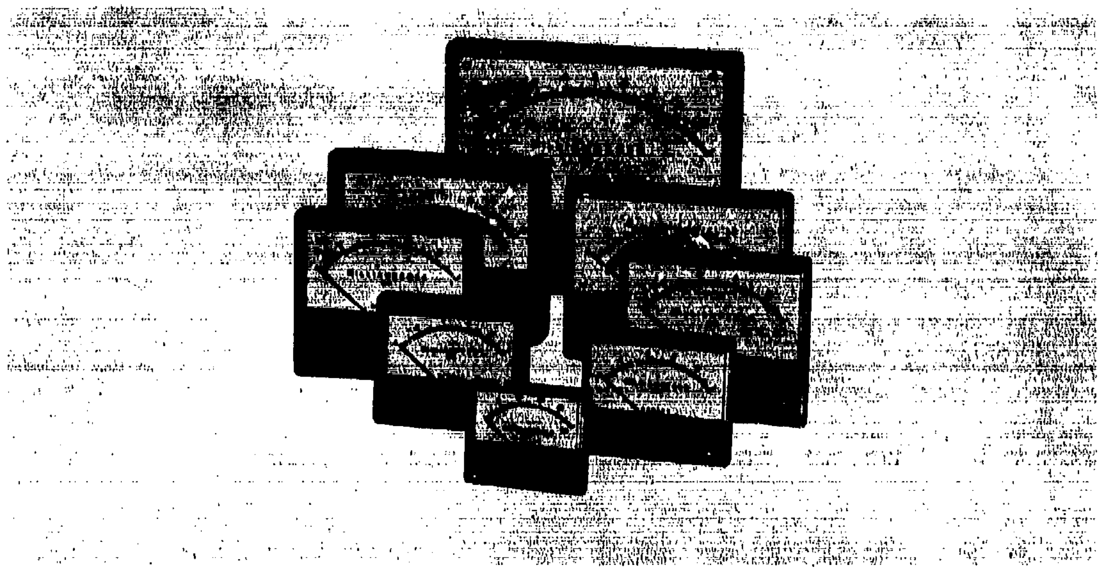
Meters designed for the measurement of specific physiological parameters by electrical means usually are scaled in the units of that parameter, i.e., pounds per square inch, degrees of temperature, etc. It must be remembered, however, that the meter measures voltage or current, and once calibrated is intended for use with transducers of specific characteristics. If transducers of different characteristics are used, the meter cannot be used without being recalibrated.

Meters are available in a variety of scale sizes, with full-scale deflections ranging from as little as 1 inch of arc to as much as 5 inches or more. The amount of panel space available for meters and the precision with which readings must be made are determining factors in selecting scale size. The scales of equipment monitoring meters found on many radio and electronic components often are no larger than 3 inches. To monitor relatively low-level physiological phenomena, however, greater sensitivity is necessary, and scales of 5 and more inches (such as are used in test equipment) are commonly employed for more accurate reading. Figure 55 shows a group of typical, commercially available panel meters.

II. Digital Meters

Digital meters accept voltage levels from other components in the monitoring system and convert them for display in numeric form. Also called in-line indicators, they offer the advantage of convenience to the operator by showing the measured quantity in a form that can be read at a glance. Digital meters also permit reading at some distance from the meter, and present a single quantity that is free of the parallax error that may occur in reading a pointer against a scale.

The digital meter consists essentially of (1) a measuring circuit, which produces a number of pulses proportional to the level of the measured signal, and (2) a digital display device, which gives a decimal reading in volts of the pulse count.



Daystrom, Inc., Newark, N.J.

Figure 55. Typical Galvanometer Panel Meters

A. The Measuring Circuit

The basic units of the measuring circuit of a digital meter are a voltage comparator circuit, reference generator, counter, and time reference oscillator (see figure 56). The comparator circuit compares the level of the unknown voltage with a reference voltage supplied by the reference generator and stops the counting cycle of the counter when the two levels are equal. The reference generator supplies the comparator with a sawtooth waveform with a controlled, uniform rate of rise and switches the counter on at the start of each sawtooth wave. The counter counts the number of pulses generated by the time reference oscillator, and the oscillator supplies pulses to the counter at a preset rate.

Briefly, the measuring circuit operates as follows: At the beginning of a sawtooth wave generated by the reference generator, the reference generator switches on the counter, which proceeds to count pulses supplied by the oscillator. The amplitude of the wave increases until it matches the amplitude of the unknown voltage in the comparator. At this point, the comparator switches the counter off. The counter thus has counted a number of pulses from the oscillator that is proportional to the unknown voltage. At the start of a new sawtooth wave, the counter is reset and switched on again, and the cycle is repeated.

For example, assume that the unknown voltage is actually 125 volts. If the reference generator is designed to produce a waveform that rises at the rate of 10 volts

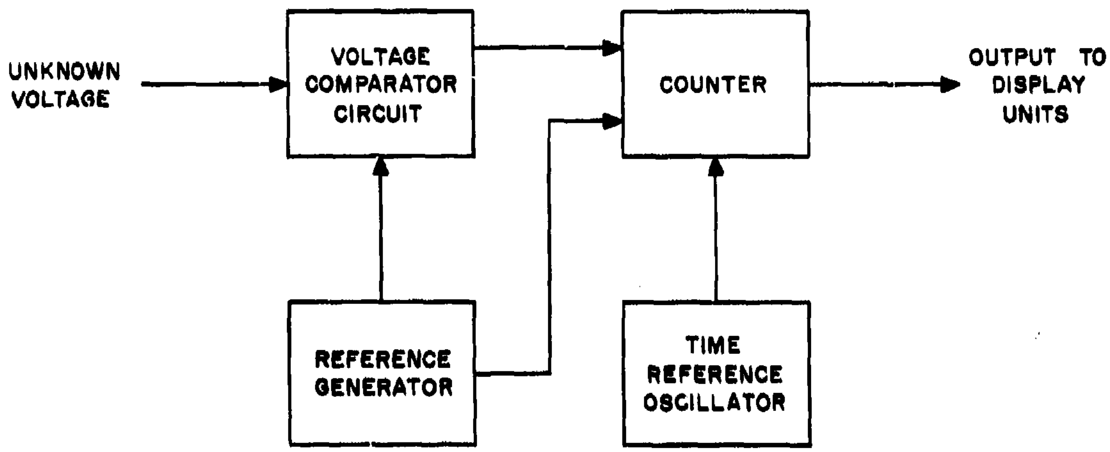


Figure 56. Basic Components of a Digital Voltmeter

per millisecond, then 12.5 milliseconds will pass during the time the waveform rises from zero to match the measured voltage. If the time reference oscillator is sending pulses to the counter at the rate of 100,000 per second, then 1250 pulses will be counted during the 12.5 milliseconds. This figure, 1250, is displayed by the meter (ref. 35).

The counting cycle is repeated as often as the reference generator cycles, which usually is at a low frequency (about 20 times a second). There are several forms of comparator circuits that can be used to stop the counting cycle. A simple one is shown in figure 57. The voltage to be measured is applied to a resistance voltage divider, and the reference waveform is applied through a diode to the junction of the divider. When the reference waveform equals the level of the measured voltage, the diode conducts, sending a pulse to the counting circuit to cut it off.

The counting circuit consists of binary, on/off devices that produce a four-line (four-bit) coded output to the display devices. Many of the display devices that provide a decimal readout cannot accept such digital inputs, so a gating or decoding stage must be provided between the counter and the display. Digital data circuits of this type are discussed in greater detail in Section VI.

B. Digital Display

A digital meter gives an in-line readout of decimal characters, so that the voltage being measured can be read directly as a number. Digital meters usually read to four significant digits, although five-digit units are available. There are numerous techniques for positioning or displaying decimal digits, including electromechanical positioning of printed or engraved characters on a wheel or drum, projection of a character onto a display screen, formation of a character by lighted line segments, and by

alternate lighting of 1 of 10 stacked characters in a group. Only the latter two techniques have wide use in digital meter devices at present. Figure 58 shows their display configurations.

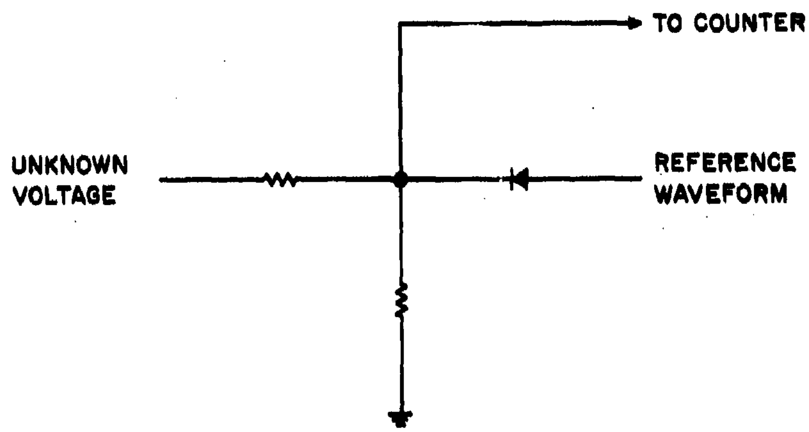
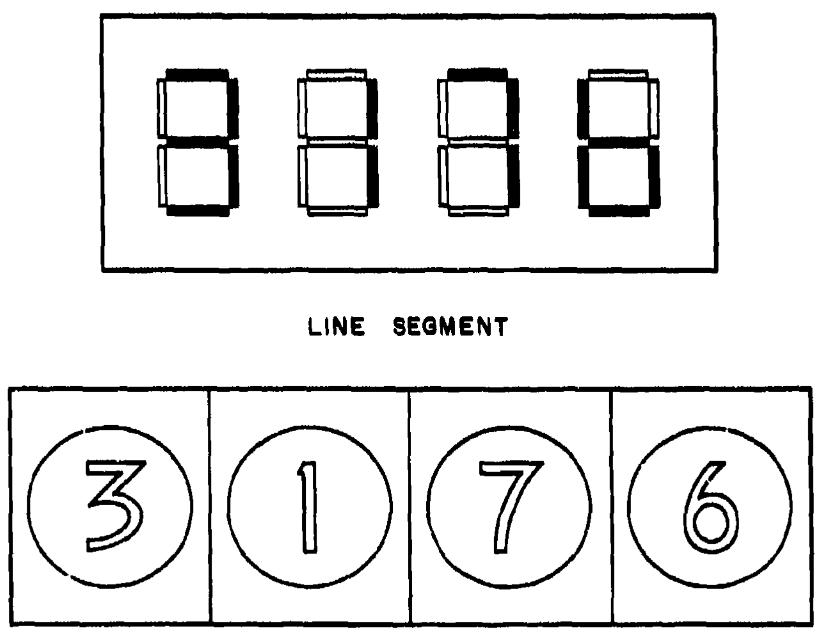


Figure 57. A Simple Voltage Comparator Circuit



LINE SEGMENT

STACKED CATHODE TUBES

Figure 58. Two Types of Digital Display Devices

These digital display devices require, either internally or as a separate input stage, a gating matrix that can accept the four-line output of the binary counting circuit described previously. This gating matrix produces electrical outputs that trigger the sequential display of decimal digits correctly when a voltage count is being made.

1. Line Segment Displays

The line segment display device (figure 58) consists of four groups of seven line segments each. In each group, sequential electrical inputs light different combinations of the seven line segments, displaying each of the 10 numerals. Several methods are used for lighting the line segments. On some meters, the line segments are made up of neon tubes or electroluminescent panels, whose circuits are energized by the signals from the gating matrix. In others, the segments are translucent strips of glass or plastic, which are lighted by small incandescent bulbs. (This latter type also is used in devices that form their characters with nine-line segment combinations, but these devices are intended primarily for the display of alphabetic as well as numeric data.)

Displays lighted by incandescent bulbs are very bright and can be easily read at some distance. The electroluminescent or neon type is less bright, and may be difficult to read if ambient light levels are high. Both types have an advantage over the stacked configurations (discussed below) in that all numerals are displayed in a single plane, which can be right at the surface of the meter panel, and there is no parallax problem; i.e., the display can be read accurately from an acute angle.

2. Stacked Character Displays

Probably the best known of the digital in-line indicator devices is the edge-lighted lucite panel. This device has been used on practically all forms of digital measuring instruments for several years. In a four-numeral in-line display, a stacked group of 10 lucite panels is used for each numeral in the display. Each panel in a group has 1 of the 10 numerals engraved on it, and each panel has a tiny incandescent bulb set in one edge; a signal from the gating matrix lights the bulb, edge-lighting the numeral in the panel.

Of course only one panel in a group is lighted at a time; nevertheless readoff can be difficult when a panel to the rear of the stack is lighted, since scattering of the light tends to light the characters in front of it as well. Also, because of the panel depth necessary to accommodate the 10-character stacks, reading the display from an angle is something of a problem.

Another unique version of the stacked character device is a multicathode display tube called the "Nixie." The Nixie is a cold-cathode gas tube employing 10

cathodes, each shaped to form a separate numeral.* The application of a potential between one of the cathodes and the anode, as triggered by the signal from the gating matrix, causes the cathode to glow, making the number formed by that cathode easily readable. The cathodes are stacked closely together near the display end of the tube, but, like the lucite panels, there is a finite depth requirement for the 10 characters, and they cannot be read accurately at too acute an angle.

GRAPHIC RECORDING DEVICES

When the actual value of a steady voltage is the significant factor in a measurement, digital display devices suffice for a monitoring system. But in the monitoring of physiological phenomena, the variations (in time) of various body functions, either their natural periodic variations or their specific responses to certain stimuli, often are of primary interest. Transient display of such transient phenomena can be obtained with an oscilloscope; but for most physiological work, permanent records of the phenomena are essential. An observer cannot evaluate a transient display of fast-changing voltages accurately, and even for the slower rates of change, permanent records permit the comparing of one measurement with another and the accumulation of quantitative data for subsequent computer processing.

One obvious way to obtain a recording is to photograph the display on a cathode ray oscilloscope (this is to be discussed later in this section on page 116). The devices discussed here are of two basic types: the galvanometric and the potentiometric (servo operated). These instruments, generally termed oscillographs, record by causing some form of stylus (or a beam of light) to traverse a strip of chart paper, leaving a visible curve which is a record of the varying input voltage.

The various recorders available are too numerous to examine in detail in this text.** The various operational characteristics of galvanometric and potentiometric records are discussed here so that the capabilities of the instruments can be weighed against the requirements of a physiological monitoring situation.

Generally, the following factors are important when evaluating graphic recording devices:

1. Frequency response - the time variation of the measured signal that can be reproduced.

2. Range or span - the amplitude variation of the measured signal that can be reproduced.

---

*Burroughs Corporation, Plainfield, New Jersey

**A recent survey (ref. 30) briefly lists characteristics on nearly one thousand recorders of various kinds, from over 200 different manufacturers.

3. Accuracy - the fidelity with which amplitude variations are recorded, which usually is expressed in some percentage of the full-scale range of the recorder, and sometimes for specific frequency ranges.

4. Form of presentation - such factors as type of stylus and paper, type of chart paper, chart speeds, etc.

I. Direct Writing Galvanometric Recorders

Basically, a galvanometric recorder is a moving-coil meter with a scribe attached to the coil assembly. The components of a galvanometric recorder may include:

1. The galvanometer movement.

2. The writing arm and stylus, which are attached to the meter arm.

3. A paper chart and chart drive mechanism.

4. Input amplifiers, usually including a separate preamplifier in addition to the driving amplifier that feeds the galvanometer (depending upon the particular manufacturer and model or type).

5. Power supplies for the amplifiers and the chart drive.

The various accessories used with this type of recorder include (1) gears and controls for changing the chart speeds, (2) additional chart markers for time and event recording, (3) chart footage indicators and rewind mechanisms, and (4) stylus accessories for forced ink supply, stylus heating, stylus vibration to overcome friction, etc.

A. Basic Operation

The basic operation of a galvanometric recorder is the same as that described for indicator operation on page 90, except that for good intermediate-frequency response, the coil and magnet must develop considerably more torque, and the spring that returns the writing arm must be much stiffer. Also, in a recorder, damping of the meter movement is especially significant.

When the pointer is deflected by the passage of current through the coil, the deflection is resisted by the spring in the coil suspension. If there were no damping, the opposing forces would cause the pointer to oscillate at the resonant frequency of the mechanical system. Damping drains off the energy producing oscillation until, with opposing forces equal, the pointer reaches an equilibrium point.

Damping is obtained in several ways. In electromagnetic damping, movement of the coil generates electromagnetic fields, either in the frame of the coil or in a

separate damping coil, that oppose the primary motion. In mechanical damping, the coil movement is immersed in a viscous medium, such as oil. In electronic damping, negative feedback in the amplifier that drives the meter movement effectively damps meter oscillations.

The effects of damping on a recording are readily seen in figure 59. For a square-wave input as shown, if the damping forces are too low, the recorded waveform shows the oscillations of figure 59(A). If the damping forces are too strong, the response time of the recorder increases, producing the lag in response shown in figure 59(C). Ideal damping is shown in figure 59(B); here, damping forces are about seven-tenths of critical. This value permits a very slight overshoot before equilibrium is reached, but faster response times are achieved than is possible with full damping.

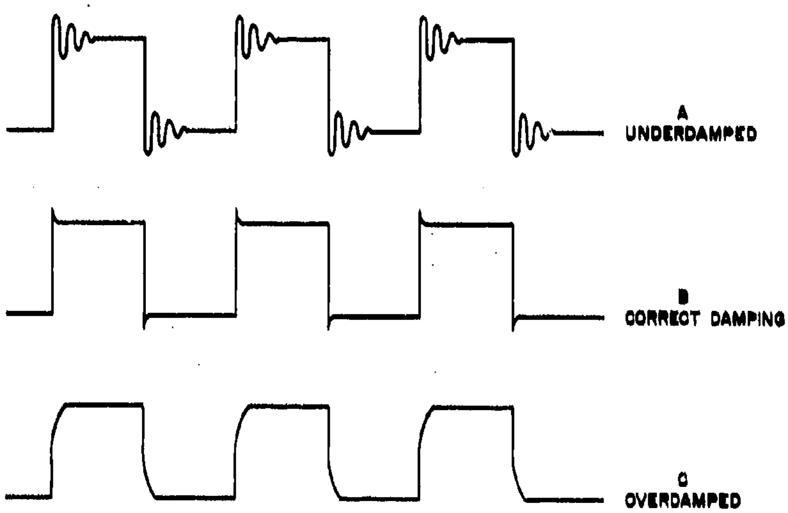


Figure 59. Effect of Damping Upon Recorded Signals

Damping forces may be adjusted in several ways. On some instruments, the pressure of the stylus against the chart paper can be changed. On others, a feedback control is provided in the driving amplifier. By whatever means is available, the adjustment is checked by applying a known waveform (such as the square wave of figure 59) to the recorder and making adjustments until the optimum recorded waveform (figure 59(B)) is obtained.

B. Frequency Response

For straightforward, direct-writing galvanometric recorders, the highest frequency that can be rec rded faithfully is about 150 cycles per second. For any instrument, the maximum response is a function of its (mechanical) resonant frequency,

and for all instruments the top frequency is limited by mechanical characteristics, such as the weight of the moving system, stylus friction, etc.

Above relatively low frequencies (about 20 cycles per second), the mechanical movement of the most sensitive meter prohibits the response from being truly linear. If a sine-wave input of about 80 cycles per second is applied to the meter, for example, the coil is deflected and the meter attempts to reach a series of equilibrium points comparable to the points along the sine curve. However, an inherent mechanical lag prohibits these equilibrium points from being reached as rapidly as the input level changes. The resultant trace, therefore, is not a true sine wave. But within the established limits for galvanometers (100–150 cps), the distortion is no worse than that which is introduced elsewhere in the system (the transducer, for example), and the near-linear response is adequate, which is reflected in the 1 percent accuracy figure that most instruments possess.

C. Range or Span

Most manufacturers of galvanometric recorders build a particular line of recording instruments around a single basic meter movement, with the fundamental full-scale response ranging from 10 microamperes to 1 milliampere. No attempt is made to provide multiple ranges in the one movement, since linearity or accuracy would have to be sacrificed. Therefore, where multiple ranges are desired, provision is made to use interchangeable preamplifiers with the meter. Thus, by using amplifier stages with different gain to match the input signal to the optimum range of the meter movement, a wide range of input signals can be graphed by the same recorder.

D. Accuracy

The accepted accuracy figure for most galvanometric recorders is about 1 percent, which is defined as an error in a given recorded level that is no greater than 1 percent of the full scale reading of the instrument. Sometimes this accuracy figure is expressed in scale division units; i.e., an error no greater than one half a division on a 50-division amplitude scale on the chart paper. The accuracy figure is seldom strictly true over the full range of the recording instrument, and manufacturers' specifications often give the appropriate qualifications. Accuracy may be quoted as less than 1 percent for 80 percent of full-scale deflection, and less than 2 percent for full scale. Or the accuracy may be specified for an optimum frequency range: one half of 1 percent below 60 cycles per second, and less than 2 percent up to 120 cycles per second.

E. Chart Drive

The recording obtained with a galvanometric recorder is a curve traced on a piece of chart paper by a stylus activated by the deflecting pointer of the galvanometer. The chart paper moves continuously during measurement, affording a record of

amplitude variations in time (time base recording). There are recorders using circular discs of chart paper that are rotated beneath the stylus, but these are used mainly in industrial type measurements. In physiological work, the strip chart is the most commonly employed, in which a roll of chart paper is passed beneath the stylus by a chart drive mechanism.

Chart drive motors may be either electric (ac or dc) or spring wound. Most recorders use a gear system that permits different chart speeds: from less than an inch to more than a foot of chart per minute (or per hour). High speed drives of as much as 10 feet per minute are possible, although the use of expanded scale measurements of this kind is unlikely in physiological monitoring. The selection of chart speeds is a function of the frequency components to be recorded, the amount of scale expansion required for the study and interpretation of the chart, and the length of time required for a given measurement or series of measurements.

In many instruments, the chart paper is moved by friction wheels. Where time base accuracy is critical, a chart drive employing sprocket wheels is preferable. The sprockets engage perforations along the edge of the chart, preventing the chart paper from slipping.

F. Type of Stylus (and Paper)

There are four principal types of mechanical writing arms used in recorders: the pen, heated stylus, pressure stylus, and electrode. Each employs its own type of chart paper.

The pen writer is among the earliest of direct-writing devices, and is still widely used. Its chief drawback is the sometimes difficulty of drawing continuous, uninterrupted traces at high-level, high-frequency signals, because of clogging of the nib with ink deposits and skipping of the pen.

The heated stylus is used with coated chart paper. The stylus is heated by electric current flowing through a resistive element, and as it travels across the chart paper, it melts off the plastic coating, exposing the black undersurface.

The pressure stylus is used with carbon-treated paper. As its name implies, simple pressure of the stylus against the writing table in the recorder leaves a black trace on the paper.

The electrode stylus is charged negative (about 100 volts) relative to a metal plate beneath the chart paper. As the stylus moves across the paper, a continuous series of discharges between stylus and plate burns marks in the chart paper.

G. Type of Curve.

The common form for graphic recordings or presentations is rectilinear; i.e., one variable such as amplitude is plotted on a scale that is perpendicular to another scale, frequently a time scale. In galvanometer operation, however, a stylus linked directly to the moving arm of the meter movement describes an arc as the arm is deflected. If rectilinear chart paper were employed with the galvanometer, displacement of the stylus with amplitude would also be seen as displacement along the time scale. Therefore, chart paper is used with divisions along the time scale that are curved to match the stylus arc. This type of recording is called curvilinear, and is commonly used with many types of pen writers. (See figure 60.) The data so recorded are accurate, but such recordings are difficult to read or interpret, and comparing different curves along the same time base is difficult.

Several manufacturers offer special linkages for connecting the writing arm of the pen recorder to the galvanometer arm, so that deflection of the pen along the amplitude axis is maintained in a line perpendicular to the time axis of the moving chart paper. This provides a true rectilinear recording. (See figure 60.)

Recorders using other stylus types, such as the heated stylus, usually provide a rectilinear recording by having the stylus contact the chart paper at a point where it is bent over a straight edge. (See figure 61.) In this type of recording, the paper is contacted not by the tip of the stylus, but by different parts of the stylus arm for different degrees of deflection; however, contact is always along the same, perpendicular (relative to the time base) line.

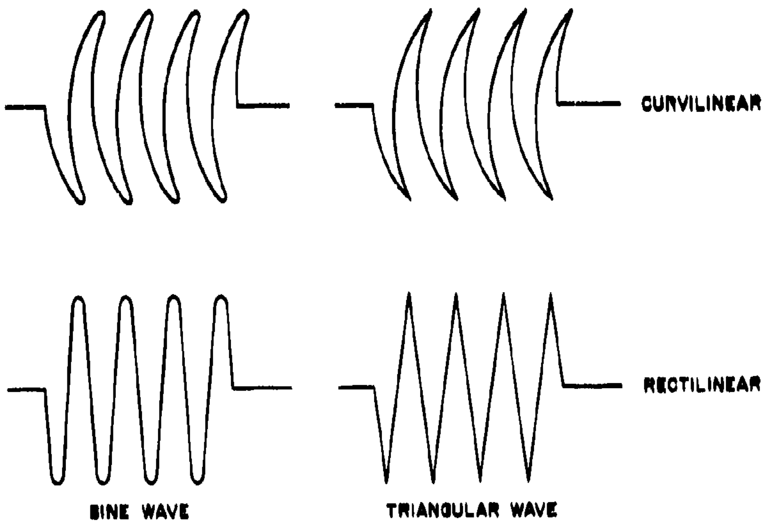
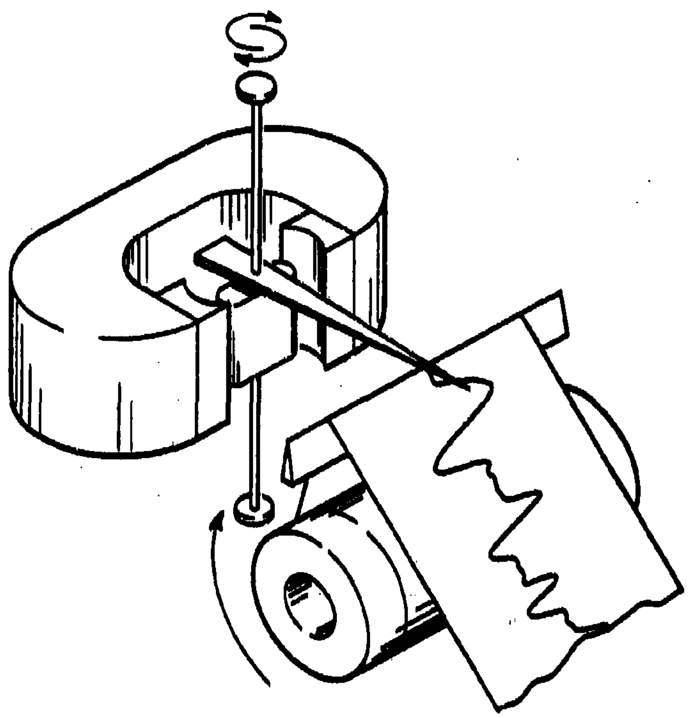


Figure 60. Curvilinear and Rectilinear Records of the Same Inputs



Sanborn Co., Waltham, Mass.

Figure 61. Rectilinear Recording With a Heated-Stylus Galvanometer

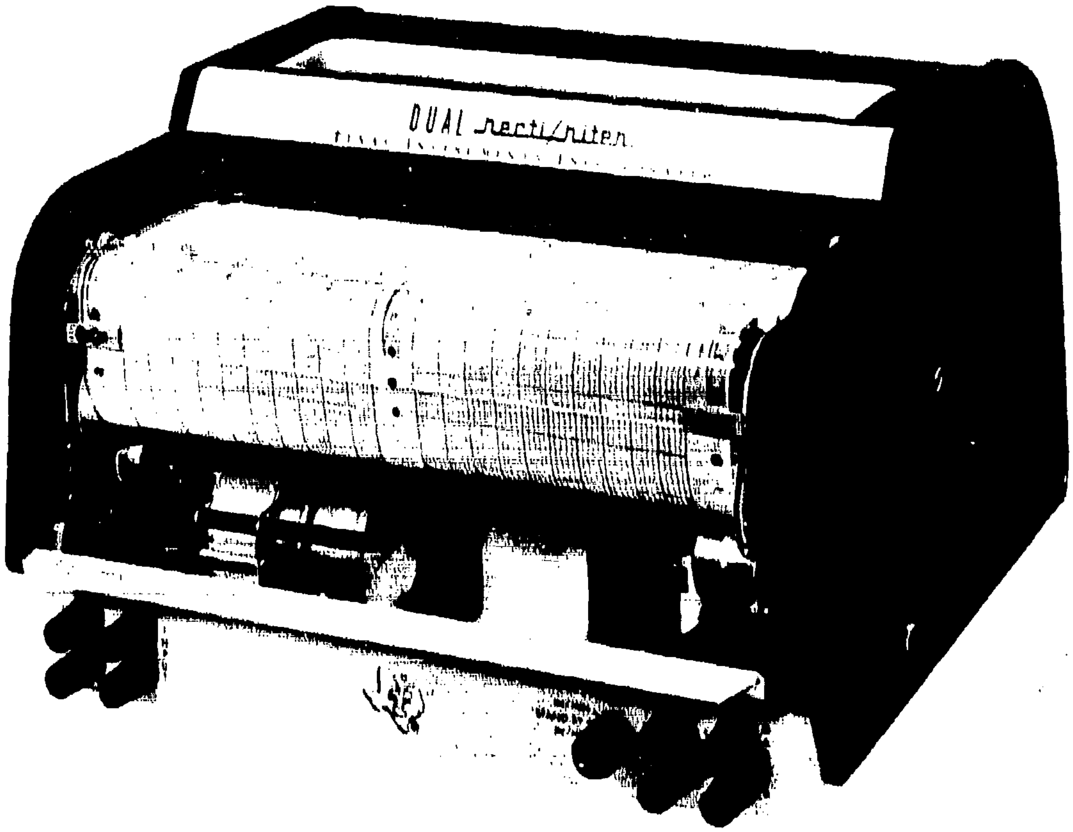
H. Multichannel Recorders

Two or more outputs of a physiological monitoring system can be recorded simultaneously with the recordings on the same chart paper. The required number of galvanometer movements are mounted in a galvanometric recorder side by side, one for each channel to be recorded. The multiple curves cannot be charted along the same amplitude axis, since the multiple deflecting writing arms would interfere with each other. However, the recordings are made on parallel charts (on one strip of chart paper) so that the same time base or time axis is shared. Galvanometer recorders with as many as 16 side-by-side charts are available.

I. Typical Galvanometric Recorders

1. All-Purpose Low Frequency Recorder

Figure 62 shows a typical galvanometric recorder that produces rectilinear ink recordings from a wide variety of voltage, current, and frequency ranges. When used with d-c millivolt inputs (likely in physiological measurements), it performs well



Texas Instruments, Inc., Houston, Texas

Figure 62. Low-Frequency Galvanometric Recorder

at low frequencies (2 to 5 cycles per second). The model shown is a two-channel recorder; a single-channel instrument also is available.*

This recorder has an input impedance of 1500 ohms, and it achieves critical damping when the source impedance (driving amplifier) is 20,000 ohms. Various milliampere ranges (with different preamplifiers) are available from 0.5 to 100. It has an accuracy of 1 percent for full-scale deflection (4-1/2 inches).

Chart drive motors may be electrical (ac or dc) or spring wound, or the chart drive shaft may be coupled to an external mechanical input. Chart speeds of 3/4, 1-1/2, 3, 6, and 12 inches per minute (or per hour) are possible. Accessories include additional timing marker pens, and a pen vibrator (with the a-c powered models)

*Texas Instruments, Inc., Houston, Texas

that provides 60-cycle excitation to the stylus, a technique that improves response at very low chart speeds and with low-level signals by reducing the friction of the stylus-to-paper contact.

2. All-Purpose Intermediate Frequency Recorder

Figure 63 shows a versatile multichannel recorder for recording at higher frequencies.* This instrument records six or eight channels of data simultaneously on the same chart paper. Each channel records on a chart with 50 amplitude divisions on a scale about 4 centimeters wide. Full-scale ranges of from 500 microvolts to 5 volts are available, depending upon the preamplifier used with the channel. The accuracy is better than 2 percent (less than one-half division error) for full-scale deflection, and better than 1 percent (less than one-quarter division error) within the center 40 divisions of the scale. The frequency response is from dc to 100 cycles per second.

This recorder uses heat-sensitive chart paper, and it is provided with individual stylus temperature control. Nine chart speeds, from 0.25 to 100 millimeters per second, are available on the multichannel instrument, and stylus temperature is changed automatically when the chart speed is changed.

3. Special Purpose Recorders

Many recorders are available that are designed for specific measurements. These devices are limited in their performance range, but considerable economy in design and construction is realized, with no great sacrifice in performance for the range of measurement for which they are intended.

Two examples, shown in figure 64, are a matched pair of dual-channel recorders designed for monitoring the physiological parameters of blood pressure (systolic and diastolic), temperature, and heart rate.** Each recorder is only 3-5/8 by 5-5/8 by 4-1/8 inches in size, and operates on 115 volts ac. Pressure-sensitive chart paper is used with a fixed chart speed of 10 inches per hour. They are low-frequency devices with a response time of 0.25 second. The accuracy of the instruments is approximately 2 percent.

II. Optical Galvanometers

A. Meter Movement

The frequency response of direct-writing galvanometers is limited by their mechanical system. The mass of the moving coil and writing arm, and the frictional

*The Sanborn Company, Waltham, Mass.
**Gulton Industries, Metuchen, New Jersey

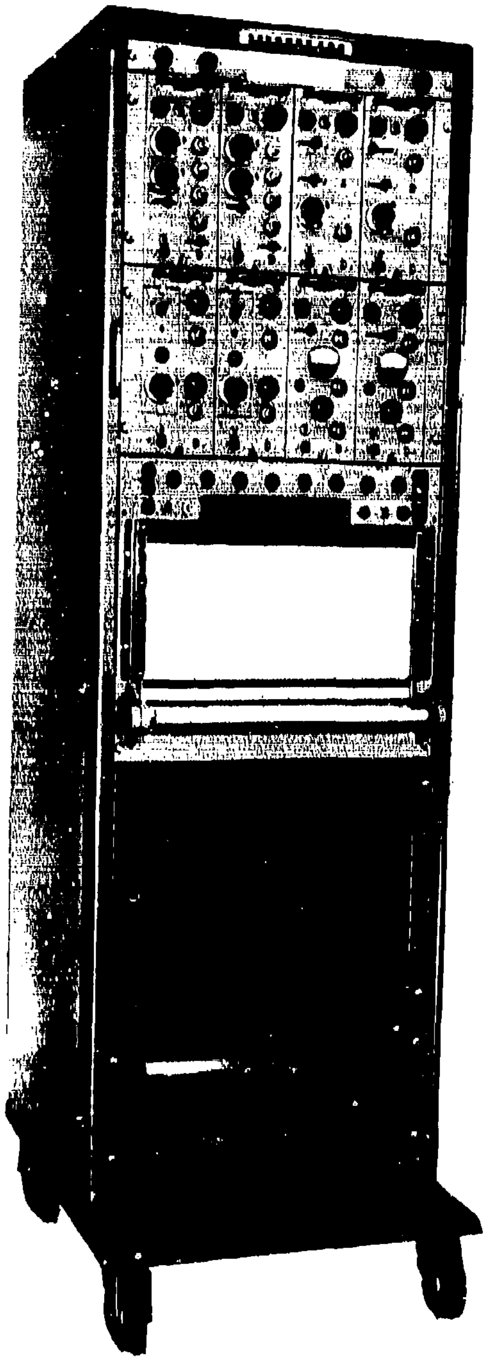
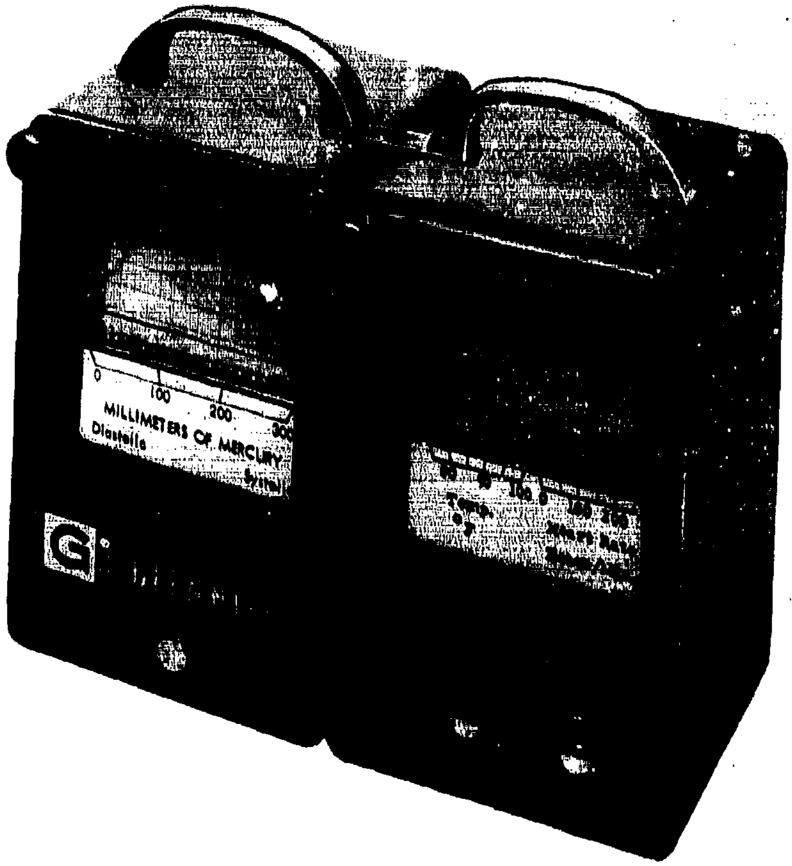


Figure 63. Intermediate Frequency Galvanometric Recorder



Gulton Industries, Inc., Metuchen, N.J.

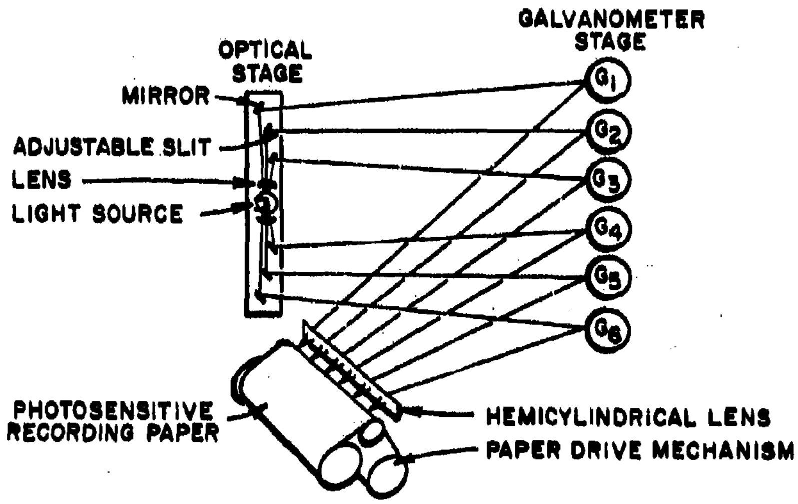
Figure 64. Typical Special-Purpose Physiological Recorders

drag of the stylus against the chart paper combine to limit the response to several hundred cycles per second. The frequency response of the optical galvanometer is in the kilocycle range.

The basic meter movement of the optical galvanometer is the same: a moving coil assembly within the field of a permanent magnet. The writing arm, however, is eliminated, and, instead, a beam of light is deflected upon light-sensitive chart paper. The only additional mass on the coil assembly is a small mirror; and, further, since the length of the light beam can be long, the angular displacement of the coil can be small; full-scale deflection is therefore possible with a small magnet and an extremely lightweight moving coil assembly. These factors result in a higher frequency response for optical galvanometers than is possible in direct-writing instruments.

B. Galvanometer Optics

The optical system in a galvanometer consists of (1) an incandescent lamp for a light source, (2) a collimating lens and aperture control to form the light into a beam and focus it on the galvanometer mirror, (3) the mirror, and (4) a lens to focus the light from the mirror into a point on the plane of the recording paper. Figure 65 shows the optical system for a typical recorder with six channels.



Sanborn Co., Waltham, Mass.

Figure 65. Optical Recording System for a Six-Channel Recorder

1. Multichannel Recording

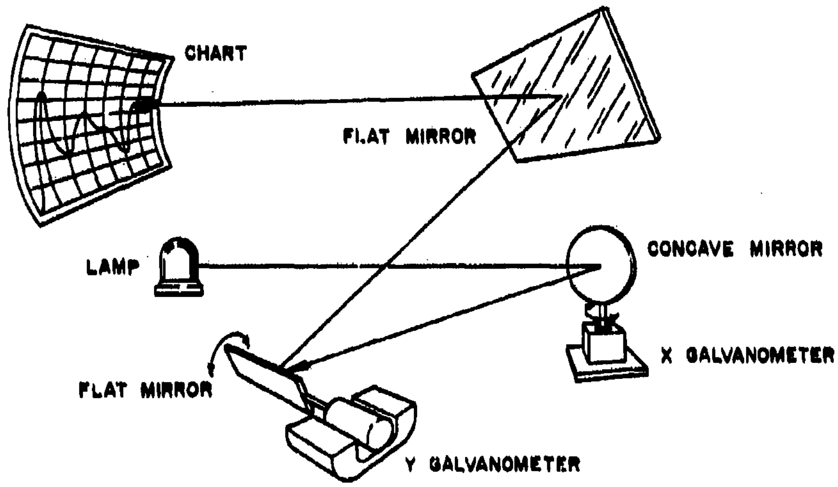
Galvanometers in commercial multichannel instruments are quite small - about the size of a pencil. Eight moving arms are mounted side by side in just a few inches of space. Recorders with 24 or more channels are available with relatively small housings. In a multichannel recording, this type of recorder has an additional advantage: since the curve is traced on the chart with light, two or more curves can cross each other without interference, and channels being recorded can be adjusted so that they occupy the full width of the chart for maximum expansion of the recorded signal.

2. X-Y Recording

All graphic recordings represent the plotting of one variable quantity against another. In most recordings, a variable is plotted against time (represented by the movement of the chart paper). In some measurements, however, it is desirable to

plot a variable against some function other than time. This is called function or X-Y recording. An example in physiological monitoring is the vectorcardiograph, in which two electrocardiographic signals, each taken from a different body location, are plotted against each other.

This type of recording can be accomplished with galvanometric recorders using an ingenious system of optics, such as shown in figure 66. One variable, X, is applied to a galvanometer with a concave mirror on its movement; the other variable, Y, is applied to a galvanometer with a flat mirror on its movement. A light beam reflected by both mirrors traces an X-versus-Y curve on a square of chart paper. (The chart paper is not moved during this recording.)



Sanborn Co., Waltham, Mass.

Figure 66. Optical System for Galvanometric X-Y Recording

C. Chart Processing

There are two principal ways to process recordings obtained with optical galvanometers. In one method, standard white photographic paper is used for the chart, and the chart-feed mechanism is treated like a camera box. The paper is loaded in a darkroom, exposure in the recorder is limited to the light beams from the galvanometers, and the exposed chart is processed in a darkroom (normal photographic procedure).

In another technique, ultraviolet-sensitive paper is used for the chart, and the light beam used in the recorder is filtered to pass only the ultraviolet light from the lamp. This paper can be loaded in daylight, and it is self-developing upon exposure to daylight. After exposure to the ultraviolet trace, the chart paper is fed out of the recorder directly, and a readable chart is available within a few seconds of exposure. This type of recorder offers the higher frequency response of the optical galvanometer

and direct access to the recording (as is possible with direct-writing instruments). The recordings do tend to fade with time, however, which would be a disadvantage in some applications.

Since the chart paper cannot be preprinted, provision must be made for tracing grid lines on the chart paper within the recorder. A separate optical system is used to trace X- and Y-axis lines (with light) upon the paper as it passes through the recorder. Various means, also optical, are used to provide time markers or to identify the various traces in a multichannel recording.

III. Potentiometric Recorders

For recording very low-frequency, low-amplitude signals, a potentiometric recorder may be employed. The basic circuit for a potentiometric (null-balance) recorder is shown in figure 67. The recording stylus is driven by a motor and linked to the moving slide on a resistance bridge that is adjusted for balance against an internal reference voltage. The signal that is measured and recorded unbalances the bridge, producing an error voltage. The error voltage is applied through a chopper and a-c amplifier to the windings of a servo motor. This motor is linked mechanically back to the moving slide, which is repositioned to a point where the error signal is zero and the bridge is once again in balance. The stylus, attached to the moving slide, traces a curve whose amplitude level corresponds to the variations in (error) voltage applied to the bridge.

Because of the mechanical limitations of the servo motor and the slide-wire stylus carriage, full-scale response of these devices is relatively slow, averaging 0.25 second or more for full-scale deflection, which limits the frequency response to about 1 cycle per second. Sensitivity is very good, however, with full-scale response for inputs of 1 or 2 millivolts, and accuracies of better than 1 percent (full scale) are easily attained. These instruments are effectively infinite impedance devices, drawing no current when in balance.

Like direct-writing galvanometric recorders, potentiometric recorders are available with charts of various widths, and with chart drives of various speeds. Multichannel

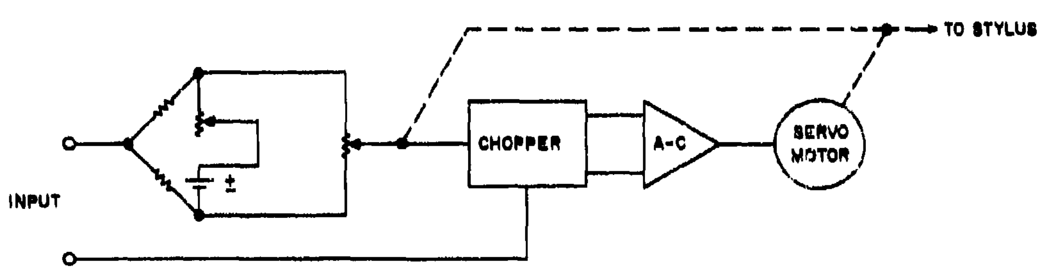


Figure 67. Basic Potentiometric Recorder Circuit

recording instruments are available with either side-by-side plots on the same chart, or overlapping plots. In the latter instance, the full width of the chart paper is used for each plot; however, there is a discrete displacement between each plot on the time axis, which is necessary so that the various styli can cross each other without interference.

X-Y or function recording is possible with potentiometric recorders in one of two ways. In one, the second variable to be plotted is applied to a second balancing circuit in which the servo motor serves as the chart drive motor, so that the rate at which the chart paper moves beneath the stylus is proportional to the second variable. In the other, the servo motor in the second balancing circuit drives the whole slide wire in a plane perpendicular to the movement of the stylus carriage along the slide wire.

CATHODE RAY OSCILLOSCOPE

One of the most useful instruments for displaying and recording high-frequency phenomena is the cathode ray oscilloscope. Since it is completely electronic in operation with no mechanical mass to move, its frequency response extends to hundreds of kilocycles, or higher, and it provides instantaneous display of complex waveforms and other rapidly occurring, transient phenomena.

The output or display of an oscilloscope is a phosphorescent trace produced on the face of a cathode ray tube when it is bombarded by a moving beam of electrons. This electron beam is the only moving part in the oscilloscope. Being virtually without mass, it can be deflected (either electrostatically or electromagnetically) at extremely high frequencies (up to many megacycles). In typical applications, the signal to be displayed is applied to the oscilloscope in such a way as to deflect the electron stream in the vertical (amplitude) plane, while an internal sweep circuit deflects the stream horizontally at a regular controlled rate. This arrangement produces the conventional time-base display. Additional inputs provide for X-Y or function display by letting a second variable (rather than the sweep circuit) control the horizontal deflection.

1. Principal Components

A. The Cathode Ray Tube

The basic unit of the oscilloscope is the cathode ray tube. (The basic construction of the tube is shown in figure 68.) A cathode ray tube is comprised of four parts: (1) an evacuated glass envelope, (2) an electron gun for producing a stream of electrons, (3) a means of deflecting the electron stream, and (4) a screen to transform the electrical energy of the electron beam into light.

The faces of cathode ray tubes vary in diameter from as small as one inch to the picture tubes used in television sets (up to 30 inches). The electron gun consists of a filament, a cathode, a control grid, a focusing anode, and an accelerating anode.

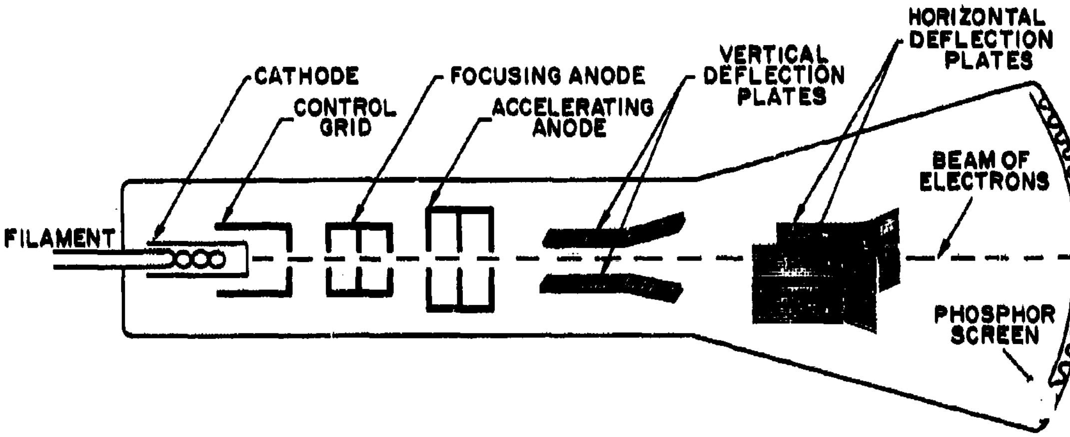


Figure 68. Construction of a Cathode Ray Tube

Electrical connections to these various electrodes are made through pins in the base of the tube, or connections through the glass wall. The cathode emits electrons, which are collimated and accelerated through the tube in the form of a beam.

The horizontal and vertical deflection plates vary the path of the beam, and a phosphor coating on the inner face of the tube emits visible light when it is bombarded with the high-speed electrons. A voltage difference between the vertical deflection plates bends the electron beam either upward or downward, depending upon the polarity of the voltage on the plates. Similarly, a voltage placed on the horizontal plates deflects the beam right or left, depending upon the voltage polarity. The beam, therefore, strikes the phosphor screen at a point determined by the voltages on the deflection plates.

B. Oscilloscope Circuitry

A block diagram of oscilloscope circuitry is shown in figure 69. The circuits consist of the following:

1. A voltage amplifier, usually a d-c amplifier, with a gain of about 10,000. This amplifier applies voltage to the vertical deflection plates.

2. A second amplifier (similar to the first) that applies voltage to the horizontal deflection plates. The input to this amplifier can be switched from the external horizontal input terminals to a sweep generator within the oscilloscope.

3. A sweep generator (with provisions for varying the frequency of the sweep) that produces a horizontal sweep on the face of the cathode ray tube.

4. A circuit, often a part of the sweep circuit, that provides a voltage pulse

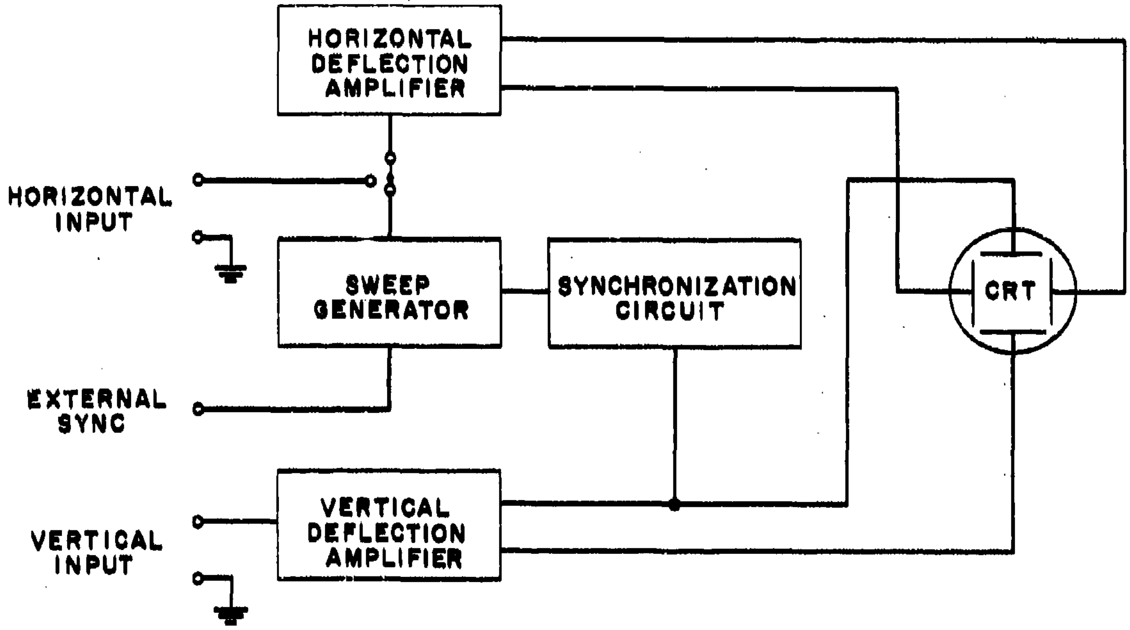


Figure 69. Basic Block Diagram of Cathode Ray Oscilloscope

during the backswing of the sweep. This pulse is used to "blank" the oscilloscope trace during the spot return.

5. A synchronization circuit that connects the vertical-amplifier output with the sweep circuit. When the frequency of the vertical amplifier is greater than the sweep frequency, this circuit causes the sweep to be driven at the higher frequency.

6. A low-voltage power supply (about 300 volts) that supplies plate voltage to the vacuum tubes used in the amplifiers, sweep circuits, etc.

7. A high-voltage power supply (about 2000 volts) that supplies the accelerating voltage placed on the electrodes of the electron gun.

Various commercially available oscilloscopes differ in the amount of stabilization of their power supplies and amplifiers, in their sensitivity and the frequency-response curves of their amplifiers, in the arrangements for increasing the rise time of their amplifiers, etc. Some oscilloscopes are made with modular plug-in components, so that the type of amplifiers or sweep circuits used can be changed easily.

Some oscilloscopes, particularly the less expensive ones, are not equipped with d-c amplifiers. Their frequency-response curves usually cut off at about 2 cycles per second at the low end. These oscilloscopes will not follow faithfully the relatively

slow changes associated with heart contraction, respiration, and other physiological stimuli.

II. Use of Oscilloscope

A. Operating Controls

A list of typical oscilloscope operating controls and their primary functions is given in table VII. In addition to operating controls, oscilloscopes often contain horizontal (X-axis) input terminals, vertical (Y-axis) input terminals, an external synch input terminal for synchronizing the sweep with an external signal, and a trigger output terminal to provide a synchronizing pulse for other oscilloscopes and recording instruments.

For ordinary usage, the intensity is adjusted to a level where the trace can be seen easily, and the focus is adjusted to make the trace as thin and as clear as possible. These two adjustments interact usually, so that when one is changed the other also must be readjusted.

TABLE VII. COMMON OSCILLOSCOPE CONTROLS AND THEIR FUNCTIONS

Control	Function
Intensity	Varies brightness of beam.
Focus	Adjusts focus of beam.
Vertical (Y) Gain	Varies vertical amplifier sensitivity.
Vertical (Y) Position	Varies initial position of beam in an up or down direction.
Horizontal (X) Gain	Varies horizontal amplifier sensitivity.
Horizontal (X) Position	Varies initial position of beam to the left or right.
Sweep Range	Varies frequency range of internal sweep generator.
Sync Selector	Selects synchronization signal from external sync terminals, 60-cps voltage, or from vertical amplifier.
Sync Level	Varies level of synchronizing pulse.

The sync selector generally is a three-position switch which is used to feed a synchronizing pulse into the time-base (sweep circuit) generator from either the vertical amplifier (internal sync position), an external voltage source (external sync position), or the 60-cycle power line (line sync position). Thus it is possible to synchronize the horizontal sweeping frequency with any of these three sources. To obtain proper synchronization, the sync level is set at zero and the sweep range is adjusted until the sweep frequency is close to the desired frequency. At this setting, the wave displayed on the oscilloscope gradually drifts to one side. The sync level then is advanced until the wave stands still, and, at this point, synchronization is completed.

B. Time-Base Display

In most applications, oscilloscopes are used with a linear time-base displacement of the horizontal axis. When the frequency of the phenomenon to be displayed is greater than 2 cycles per second, the time-base generator in the oscilloscope is adequate. However, if the signal frequency of interest is lower than 2 cycles per second, such as an electrocardiogram or the pressure pulsations in a blood vessel, a complete pattern or cycle cannot be obtained on a single sweep, without resorting to special procedures. A linear sweep must be obtained that is slow enough to display two or three complete cycles of a signal having a frequency of less than 2 cycles per second.

The simplest method for obtaining a slow speed is to set the sweep range control for external capacitance, and connect a capacitor between the horizontal input terminal and ground. This external capacitor, which should range in value between 2 and 20 microfarads, then controls the sweep frequency. The leakage resistance of most electrolytic capacitors is too great to permit a linear sweep. As a result, more expensive mica or oil-filled capacitors should be used.

C. X-Y or Function Plotting

If an input other than the time-linear voltage of the sweep generator is used in the horizontal input of an oscilloscope, the waveform displayed will be a function of the signals applied to both the Y and the X axes (two variables). For example, if a signal representing the length of a muscle is fed to the X axis and a signal representing the muscle's tension is applied to the Y axis, the result will be a length-tension diagram complete in all details. Vector electrocardiographs are obtained by driving the Y axis of an oscilloscope with the electrocardiogram sampled from electrodes placed on one axis of the body and driving the X axis with an electrocardiogram simultaneously sampled on another axis of the body.

D. Multichannel Operation

Cathode ray oscilloscopes ordinarily are single-channel devices, and it often is necessary to observe or record several phenomena simultaneously. An expensive and

not too often used remedy to this problem is to use a multigun tube. The multibeam units are built so that they can be used together with a single sweep generator, or they can be used separately. The available definition and brightness of the traces produced by a multigun tube are excellent; however, the number of available display channels is limited to the number of guns contained in the tube.

If the traces do not have to be physically close together for comparison, two or more single-beam oscilloscopes may be set side by side, with each additional oscilloscope synchronized by the sync output trigger of the first. A simple and relatively expensive solution to the problem of producing a multichannel display is to use an electronic switch. When the switched input method is used, a number of inputs are electronically connected in sequence to the horizontal deflection plates of a single-gun tube. Each input is exchanged for the next at the switching frequency so that samples of each input are obtained at a frequency equal to the switching frequency divided by the number of channels. The highest possible sampling frequency should be used to provide a large number of short samples rather than a small number of long samples. The maximum usable switching frequency usually is a property of the electronic switch.

III. Recording With the Cathode Ray Oscilloscope

The phenomena displayed on an oscilloscope can be recorded on film in three ways: (1) making still pictures of the oscilloscope face, (2) taking moving pictures, and (3) using a kymograph camera.

When still pictures are made, the oscilloscope's time base generator is used to produce a horizontal sweep, and the face of the oscilloscope is photographed at desired intervals. Although this method is inexpensive and easy to set up, still pictures do not provide a continuous record of data versus time.

Motion picture photography can be used to record the standard oscillographic display. However, the record is on film and can only be displayed by projection. It does not provide a convenient record that can be studied and used for measurements.

A kymograph camera is a device used to convert a cathode-ray tube display into a permanent record of data versus time on recording film. When a kymograph camera is used, the horizontal sweep of the oscilloscope display is eliminated, leaving only vertical deflections of the electron beam which are caused by the stimulus being monitored. A strip of film then is continuously pulled past the face of the oscilloscope to generate a time base. The result is a permanent and continuous record of the cathode-ray tube display.

AUDIO-VISUAL PRESENTATION DEVICES

Section I included a brief discussion of the problem of monitoring audio, visual, and environmental data in conjunction with physiological data. While it is not within the scope of this handbook to treat fully the subject of sound and video systems, the following paragraphs will cite briefly the equipments required for the display and recording of voice communications and visual observations. (The presentation of environmental data, by and large, can be accomplished with the devices already described in this section.)

I. Audio Devices

The basic reproducer or transducer in an audio or sound system is the loudspeaker. For simple voice communications, requirements for the reproducer are not critical, and speakers with frequency responses no higher than a few kilocycles will suffice. In a reasonably well-appointed laboratory or telemetry station, an existing public-address system may be conveniently connected into the monitoring system to reproduce voice communications.

If a record of the voice communications is desired, a modest audio tape recorder is adequate. This will of course be separate from the tape devices used to record physiological data. Disc recorders of the office-dictation type can also be adopted for recording voice communication during physiological monitoring, but they are not as suitable as tape, either in terms of long-time recording or quick retrieval of selected parts of a recording.

II. Visual Presentation Devices

Simultaneous visual observation of a monitoring program entails the use of a closed-circuit television pickup and display system. Television display devices are comparable to the familiar broadcast television receiver used in the home. Broadcast receivers can in fact be adapted to use in a closed-circuit system, where low-power radio transmission at broadcast frequencies will not interfere with commercial broadcasting.

More common, however, are video monitors designed especially for closed-circuit display. Video monitors contain only the video circuits and kinescope display tube; the radio-frequency circuits associated with transmission of the picture signal are either unnecessary or are included in other components in the system. Monitors are available ranging in complexity and expense from high-resolution broadcast-studio types to relatively cheap industrial television units, where picture quality has been compromised for simplicity, ruggedness, and lower cost.

Recording of visual data from a television system is done in two ways. The simplest is photographing the display of a television monitor with a motion-picture camera. This will entail some loss of picture quality between the initial television display and the projected picture obtained with the film, even under the best conditions of filming. The other method, video tape recording, will preserve an electronic signal that guarantees a video display of as good quality as the video system was capable of obtaining initially. But video tape recorders are complex, expensive instruments, and their use would be inconsistent in any but the most sophisticated monitoring system.

Section IV

SPECIAL SYSTEM DEVICES

MAGNETIC TAPE RECORDING

1. Application to the Measurement Situation

Magnetic tape recording provides a means for holding or storing experimental data in a compact, readily available form. In the overall measurement situation, magnetic recording is an intermediate process that is used at the same point in a monitoring system as the data transmission or telemetry link. Data therefore are recorded in their original, electrical form and, for interpretation, must be converted to visual form by feeding the recording into the customary graphic display and recording equipment.

The greatest value of magnetic tape recording is that it can capture experimental data in real time (during the actual measurement) with a minimum of instrumentation. As a result, there is great flexibility in the subsequent display, interpretation, or other processing of the data. If all these steps were to be accomplished in real time, each measurement would have to be planned carefully in advance and the instrumentation required would be necessarily more complex. For example, tape recorders can be packaged in compact, rugged units for use in extreme environments (excepting severe angular accelerations) at remote locations. So used, they can either (1) eliminate the need for telemetry, or (2) permit telemetry transmission at some time later than the actual measurement.

More important, perhaps, is the capability of repeating the experiment as often as desired once the experimental data are recorded on tape. Repeated visual or aural display is permitted, as are repeated computer processing operations. Further, by using different tape speeds for reproducing than were used for recording, time bases can be readily changed: hours of recording can be compressed into minutes, or high-speed phenomena can be reproduced at slower speeds for more detailed study.

There are three basic magnetic recording techniques suitable for recording physiological data: (1) direct, (2) indirect (or FM subcarrier), and (3) digital (or pulse code). The components required for recording and playback are noted below (see figure 70):

1. Signal conditioners, both input and output.
2. Magnetic heads for recording and reproduction.
3. Magnetic tape, the recording medium.
4. The tape transport, to move the tape past the heads at a controlled rate.

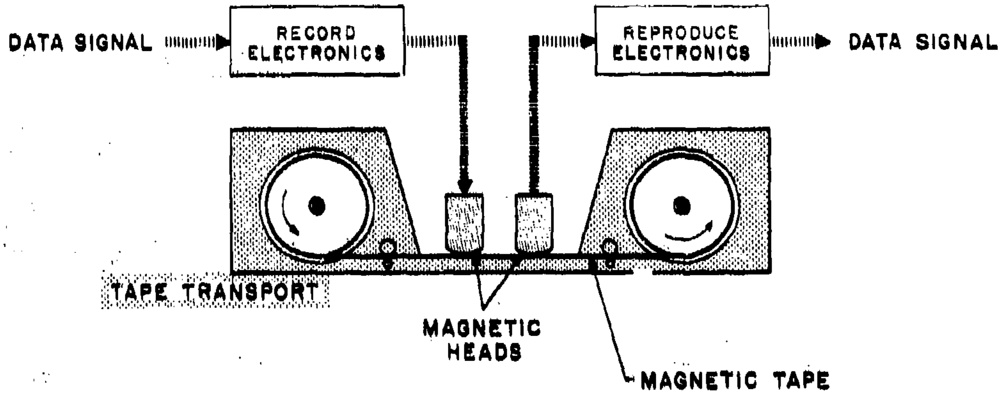


Figure 70. Components for Magnetic Tape Recording

Direct recording is recording in the magnitude domain, and it has two major drawbacks. First, steady or slowly changing signal levels cannot be recorded by this process; second, precise recording of amplitude variations is always subject to distortion because of certain uncontrollable characteristics in the magnetic materials used in magnetic tapes. For these reasons, indirect (FM) and digital recording techniques, working in the time domain, are more often used when highly accurate recordings (including dc and low frequency) are required.

Following a discussion of basic recording principles, the various recording techniques are described and component factors are noted in further detail.

II. Basic Recording Principles

In all types of magnetic recording, the signal to be recorded generates and controls an electromagnetic flux field, and this flux field changes the arrangement of magnetic material in the tape permanently or semipermanently. (See figure 71.) The signal to be recorded is applied to the coils of a circular electromagnet whose poles are separated by a narrow gap. When signal current flows through the coil, a field of magnetic flux lines is generated in the magnet which crosses the gap in the manner shown.

Magnetic recording occurs at this gap when the recording medium is drawn through the flux field. The medium is a plastic tape with a thin coating of magnetic material on one side, usually iron oxide. Before exposure to the flux, the magnetic material essentially is neutral; when passing the gap, it is magnetized to a degree proportional to the amount of flux present, and this magnetization is retained as the record.

The magnetic flux at the gap is directly proportional to the current signal in the coil: i.e., $\phi = KNI$, where I and N are the current and the number of turns in the coil, and K is a constant. The residual magnetism left on the tape, however, after it is removed from the field is always less than ϕ, and it is not linearly proportional to ϕ;

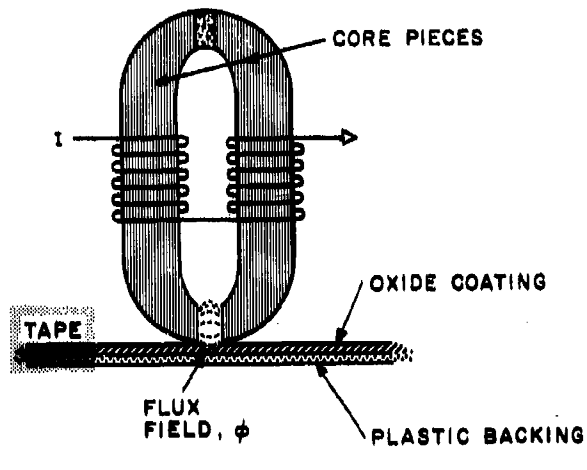


Figure 71. Magnetic Recording Head Arrangement

this is because of hysteresis, a nonlinear property of magnetic materials. Thus, for a change in flux from extreme negative to extreme positive, the residual magnetism, M_R, transferred to a tape is approximately as shown in figure 72. The curve in figure 72 actually is an average curve, because with the phenomenon of hysteresis, there is a definite range of values for M_R obtainable for any one value of ϕ. Also, it might be noted that the curve is linear over those values of ϕ somewhat removed from zero and below (or above) the plateaus of positive (and negative) saturation. By using a high-frequency bias signal to maintain the flux field somewhere between zero and saturation, a linear recording of slow-varying input signals can be obtained (ref. 58).

Recording normally is a time-based or amplitude-versus-time function, and in magnetic recording the time base is derived from the ordered movement of the tape through the magnetic field at the gap. The residual magnetism left on any portion of the tape is an absolute value depending upon the level of the flux across the gap at the time that the tape is about to move out of the gap field. The speed of tape movement is critical to the accurate recording of a varying intelligence signal and must be closely related to the rate of change or frequency of the intelligence signal. There is a maximum amount of intelligence (number of flux changes) that can be impressed upon a tape without overlapping. The maximum of flux reversals that can be recorded on a tape is about 4000 per inch, or 2000 cycles per inch (ref. 48). This means that to record a signal with 30,000-cps frequency components, the tape would have to move past the gap at a speed of 15 inches per second.

When reproducing a magnetic recording, not just the tape speed but the width of the gap as well must be closely related to the rate of change or frequency of the

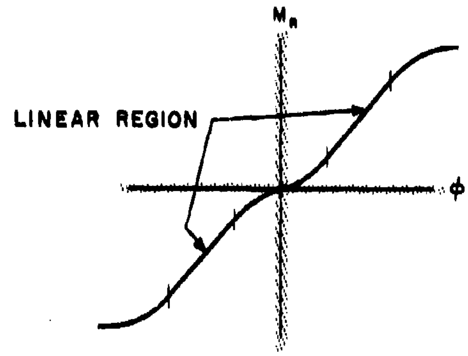


Figure 72. Residual Magnetism (M_R) Produced by Flux (ϕ)

recorded signal. The reproduce head is passive (not active like the record head). It does not change the residual magnetism on the tape, but does respond to the average flux value of that portion of the tape present at the gap in the circular magnet in the reproduce head. Thus, the output voltage that is generated in the coils of the magnet is determined not by the magnitude of the flux field, but by the time rate of the change of the flux. That is,

$$E_{out} = KN \frac{d\phi}{dt}$$

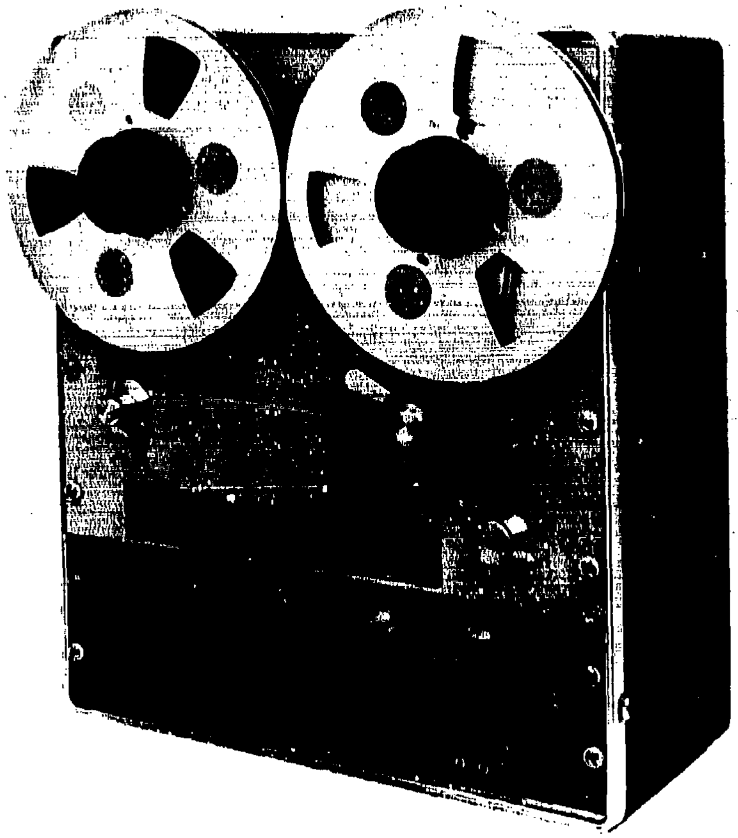
or the output voltage is proportional to the derivative of the flux within the magnetic gap.

Therefore, if there is no change in the magnitude of the flux (no change in the recorded signal level), there is no output. Similarly, because this is an averaging phenomenon, the gap of the reproduce head must be kept quite narrow with respect to wavelength. If the gap is too wide (approaching one wavelength), offsetting flux changes, such as the positive and negative swings of a sinusoid, will not be reproduced, since their average value over the region in which they are recorded is zero.

For a constant-current recording, this type of frequency-sensitive reproduction also causes the output voltage from the reproduce head to increase directly with frequency. This effect can be compensated for by applying the head output to an amplifier, called an equalization circuit, with inverse characteristics; i.e., the amplifier output decreases as the input frequency increases.

As with most other signal-handling stages in a system, the magnetic tape recorder is designed with a relatively high input impedance of about 10,000 ohms and a low output impedance of about 100 ohms. It produces normal recording levels with an input signal voltage of about 1 volt, and it delivers an output at about the same level. The frequency response is a function of the tape speed. For a portable instrumentation

recorder such as shown in figure 73, the direct recording frequency response may vary from 50 to 5000 cps at 1-7/8 inches per second to 50 to 40,000 cps at 15 inches per second.



Ampex Corp., Redwood City, Calif.

Figure 73. Typical Four-Channel Recorder

III. Magnetic Recording Techniques

A. Direct Recording

The basic recording principles just described are fundamental to all magnetic recording, and different techniques are but adaptations. A typical direct-recording system is shown in figure 74.

The changing data voltage is converted to a changing current at the input amplifier at a level high enough to operate the recording head. The bias oscillator

generates a high-frequency bias current (about 100,000 cycles per second) to ensure linear recording, and the two currents are added in the bias filter and applied to the coil in the recording head.

The recording head converts the changing current into varying magnetic flux, changing the residual magnetism on the magnetic tape as it moves past the recording head. This magnetism on the tape, in turn, produces a change in the output voltage from the reproducing head that is proportional to the rate of change of flux in the reproduce head.

This changing voltage is preamplified and applied to an equalization amplifier to linearize the amplitude-frequency response. Part of the output should be fed back to a phase-equalizing circuit if frequency- or phase-modulated subcarriers are being recorded to compensate for nonlinear phase response near the limits of the recording passband. The signal is finally applied to an output amplifier to be raised to a level adequate to drive visual display or graphic recording devices.

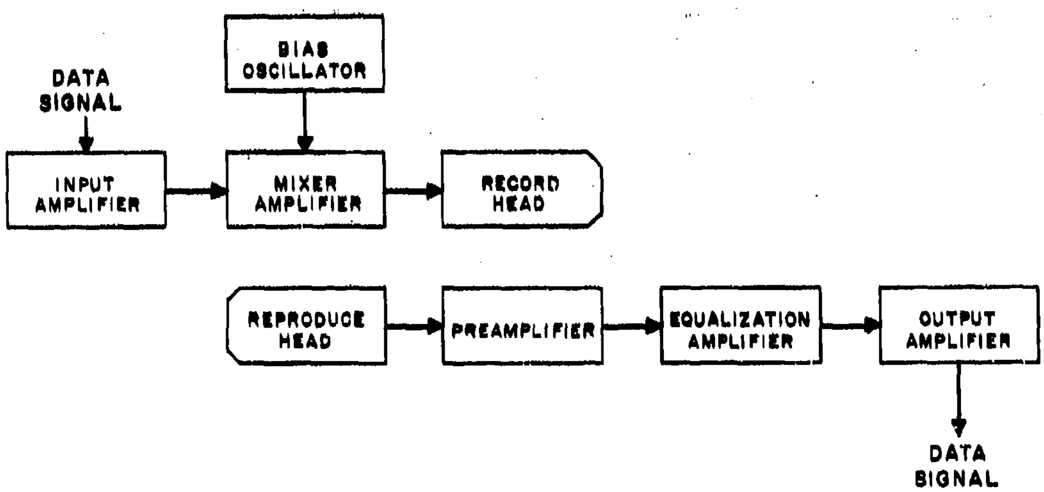


Figure 74. Direct Recording Electronics

B. Indirect Recording

One of the drawbacks of the direct-recording process is that very low frequencies cannot be recorded. A much used instrumentation technique to overcome this limitation is called frequency-modulation recording. In this method, the signal to be recorded is applied to a modulator where it frequency modulates a higher frequency carrier, and the modulated carrier then is applied as an input to a direct-recording device. On playback, the output from the reproduce electronics is demodulated and filtered to remove the carrier frequency. (Refer to section V for a discussion of frequency modulation.)

This technique permits recording down to dc (0 cycle per second). However, other requirements governing FM recording, notably the limit on the maximum deviation of the carrier frequency (normally 40 percent), reduce the overall bandwidth or frequency response of FM recording as compared to the frequency response of the same instruments for direct recording. For example, the portable instrument shown in figure 73 when used for FM recording has a frequency response that varies from 0 to only 312 cps at 1-7/8 inches per second, and from 0 to 2500 cps at 15 inches per second. Higher frequency response is attainable; larger and more versatile equipment, with tape speeds of 60 inches per second, have responses up to 20,000 cycles per second.

C. Digital Recording

Magnetic recording of data that has been encoded in pulse format is most often used with binary digital coded PCM data. Such data may be recorded serially on one tape track or in parallel, using several tracks, one for each digit in the code. (Refer to section VI for a discussion of these digital data techniques.)

Digital recording is designed about the binary coding format. In binary form, all data is coded as + or -, 1 or 0. For recording, therefore, the analogous states are two levels of magnetic flux saturation: such as full saturation first in one direction and then the other, or full saturation and neutral (or some predetermined bias level). There are three basic techniques for binary coding (see figure 75):

Return-to-Zero (RZ). In this method the tape normally is magnetically neutral, and each pulse or bit in the code produces saturation in one direction or the other, depending on whether the bit is 1 or 0.

Non-Return-to-Zero (NRZ). In this method the tape is always saturated in one direction or the other, and the polarity changes each time a bit occurs that is different in polarity from the preceding bit. A variant of this method has the polarity change each time a 1 bit occurs.

Return-to-Bias (RB). In this technique the tape is normally at a predetermined (biased) flux level, and the flux swings positive each time a 1 bit appears (and then returns to the bias level).

Signal conditioning for all these techniques entails driving the recording head to saturation, with the output of such bistable pulse circuits as flip-flops. For reproduction, the differential output from the reproduce head must ultimately drive another bistable device to restore the binary pulse train.

1. Serial Recording

Except for the special signal conditioning required for saturation recording, serial recording requires no special tape recording techniques. A direct recording

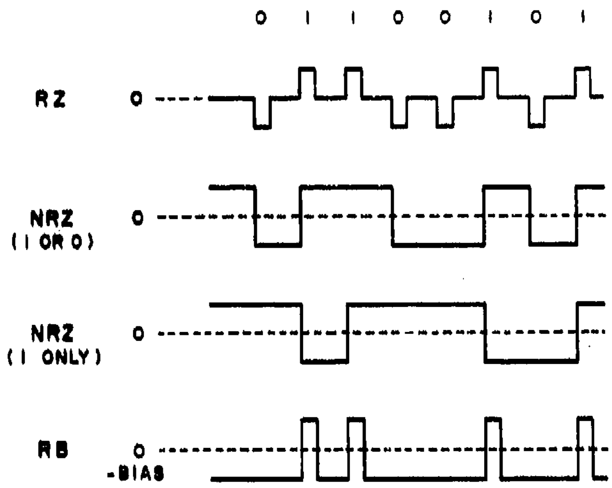


Figure 75. Pulse Trains for Typical Binary Value (01100101)

system may be employed, with the only requirement being that the upper frequency limit of the system is at least one half of the pulse per second information rate. Thus if the pulse or bit rate is 60,000 per second, a system with a 30,000-cps range could handle it, using NRZ recording. (The 30,000-cps rate means a flux reversal rate of 60,000 per second.)

Obviously, serial recording of PCM data makes full use of the information storing capability of magnetic tape. Further, such recording is extremely accurate, being relatively free from data loss or distortion caused by tape speed variations. One possible source of error in serial recording is tape dropout. Tape dropout is the loss of flux resulting when an imperfection in the surface of the tape displaces the tape from the record or reproduce head enough to permit loss of flux contact between head and tape. If the pulse rate (pulses per inch of tape) is high enough, dropout could mean the loss of one or more significant bits in a binary code, producing a wrong data signal. Reliability can be increased by redundant recording of the PCM data on two parallel tracks.

2. Parallel Recording

In parallel recording of PCM data, each digit in the binary code is handled in a separate channel, so that a whole code group or binary word is recorded simultaneously across the width of the tape, rather than in sequence along the length of the tape, as in direct, serial recording. Therefore, as many tape tracks are required as there are significant bits in the binary code (plus additional tracks for clock pulses and parity check pulses) resulting in relatively poor tape utilization. Parallel recording is an extremely useful technique, however, with digital computer operations.

Like serial recording, parallel recording of PCM data has an inherent capability of high accuracy, being little affected by tape speed variations. Error may still be introduced by tape dropout and by tape skewing. Tape skew refers to the longitudinal shifting of the parallel tracks on the tape because of changes in tape transport tension. This shifting can cause loss of the proper time relationship between the parallel tracks. Tape skew is probably the greatest factor limiting the pulse rate that can be recorded without error.

D. Predetection Recording

Another useful technique in magnetic tape recording is the recording of telemetry data before the removal of the r-f carrier. Sometimes called predetection recording, it puts telemetry data on tape for storage in essentially transmitted form -- usually after conversion in a receiver to a lower, intermediate frequency, but before detection and removal of all carrier components, as opposed to recording of r-f detected subcarrier frequencies (ref. 48).

Predetection recording has only been possible since the introduction of recording systems with extremely wide bandwidth (1 million cycles a second and higher). There are two basic systems used for predetection recording. One, the longitudinal system, uses equipment much like that employed for direct or FM recording; the increased bandwidth (up to 1 megacycle) is achieved by increasing the tape speed to 120 inches per second and by increasing the longitudinal resolution of the recorder (by reducing the width of the gap in the reproduce heads).

For recording, the telemetry data are taken from the output of the intermediate-frequency stage of a receiver (at about 5 megacycles) and heterodyned down to about 1 megacycle and recorded. For the playback of the recorded data, the signal is taken from the reproduce circuits of the recorder and mixed with the same oscillator frequency used for recording to restore the 5-megacycle intermediate frequency. The data are then fed back into the receiver for subsequent carrier and subcarrier detection and channel separation.

The second predetection system uses equipment similar to that first developed for video tape recording. Known as transverse recording, the magnetic heads rotate to scan the tape laterally as it moves past the heads longitudinally. By this technique, head-to-tape speeds in excess of 1000 inches per second can be attained, and the longitudinal tape speed can be adjusted to yield proper track spacing between the lateral scans. Bandwidths up to 5 megacycles are attainable, so that intermediate frequencies from a receiver may be recorded directly without further conversion to a lower frequency.

IV. Component Factors

A. Magnetic Heads

Figure 76 shows a typical seven-track magnetic head. Each channel is served by an identical core piece, which is composed of two core halves of material with high magnetic permeability. Both halves are wound with an identical number of turns of wire to form the electromagnetic coil. More turns of finer wire generally are used on the reproduce head, where the output from the head is directly proportional to the number of turns. The two halves are separated by two gaps, which are filled with non-magnetic separators. One gap contacts the magnetic tape under the head, where recording and reproducing occur; the other gap is designed to maintain magnetic symmetry of the core halves. In a record head, the gap width is fairly standardized at about 1/2 mil; in a reproduce head, the gap normally is much narrower (about 1/12 mil) (ref. 58).

The cores for separate channels are separated by Mu-metal shields to prevent intercoupling between adjacent cores and coils. The standard spacings for magnetic recording are as follows: each track on the tape, 50 mils; center-to-center track spacing, 70 mils; and center-to-center head core spacing, 140 mils. Therefore, to record seven 50-mil tracks on a 1/2-inch tape, two head stacks must be used on the record head, with channels or tracks interleaves as indicated in figure 77. This arrangement naturally produces an offset between the two stacks, and this distance must be maintained uniform to establish the relative timing between the tracks recorded on the separate stacks. Thus, for the simultaneous playback of all seven channels shown in figure 77, with a distance between the stacks of 1.5 inches and a tape speed of 15 inches per second, the data recorded on tracks 1, 3, 5, and 7 will precede the data on tracks 2, 4, and 6 by exactly 1/10 second.

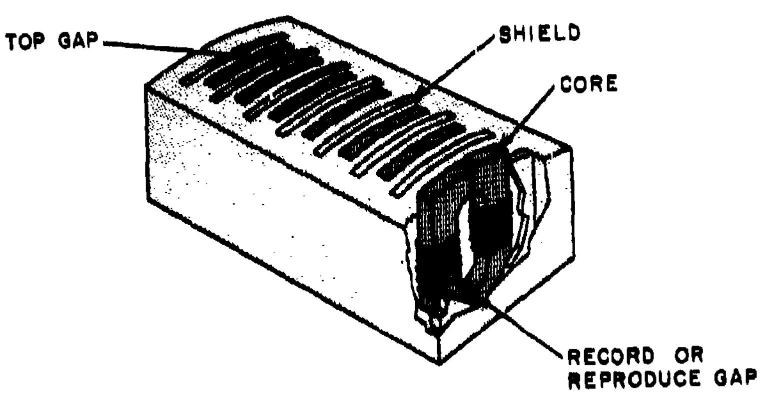


Figure 76. Seven-Channel Magnetic Head (Single Stack)

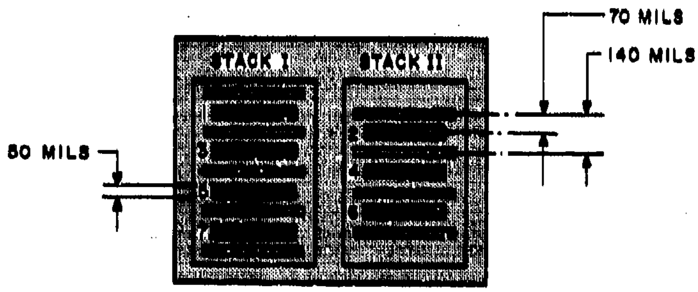


Figure 77. Seven-Channel Magnetic Head (Two Stacks, Interleaved)

Instead of using double-stacked recording heads, adequate intertrack spacing is possible simply by employing wider tape. With a 1-inch wide tape, seven tracks can be recorded without interleaving, using a recording head with a single stack of seven magnetic cores.

Double-stack recording heads for the digital recording of data cannot use interleaved tracks, because it is essential that the digital data in each channel making up the digital code be recorded in parallel (including the redundant channels and the parity check channel). However, because the dynamic range is not so great and the channel crosstalk problem is proportionally smaller, the cores can be spaced much closer together for digital recording. Therefore, as many as sixteen tracks can be recorded on 1-inch tape with a single-stack head.

After proper spacing, probably the next most important feature in the construction of tape heads is the finish of the surface that contacts the tape for recording and reproduction. The general contour is a gentle curve to minimize interaction between the magnetic core elements and portions of the tape which do not lie within the gap region. This surface must be polished to a very smooth finish so that uniform, intimate contact is maintained with the tape, and tape dropout is minimized.

B. Tape Transports

The tape transport is the mechanism that moves the tape past the magnetic heads of the magnetic recorder. The most important consideration in the design of a tape transport is constant tape speed -- smooth, linear motion of the tape past the magnetic heads. The principal parts of the transport are shown in figure 78.

The magnetic tape is wound on the takeup reel which is driven by a motor during recording or reproduction. The supply reel is used for storing the tape when it is not in use on the recorder. (This reel also has a motor for rewinding the tape after recording or reproduction.) Both reels have brakes to decelerate the reel smoothly and quickly when the driving motor is turned off.

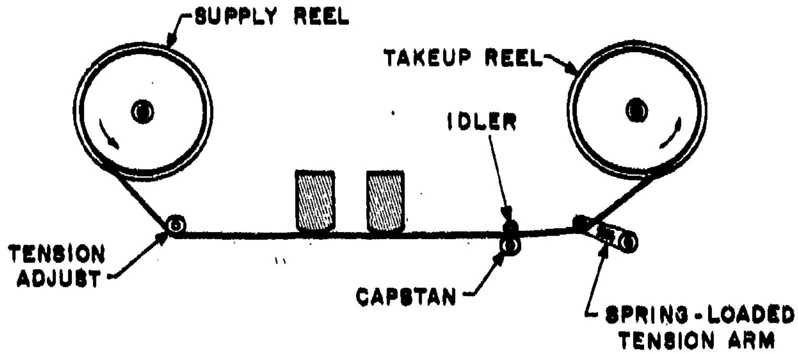


Figure 78. Basic Tape Transport Mechanism

The capstan and idler control the speed of tape motion. The capstan is driven by a constant-speed motor, and it supplies friction drive to the tape, which is pressed against the capstan by the spring-loaded idler.

Accessories to the basic drive include the tape tension arm, which is spring-loaded to take up any slack tape between the capstan and the takeup reel, and the inertia wheel, which smooths out variations in tape speed resulting from irregular motion of the supply reel.

Variations in tape speed are called flutter or wow. Since the reproduction process in magnetic recording is a frequency-sensitive phenomenon, these speed variations are obviously objectionable, because they appear to the reproduce head as changes in the frequency of the recorded signal. Flutter and wow therefore appear as noise in an FM recording, as errors in pulse recording, and as a general distortion of the time base in any type of recording.

Speed variations can be caused by many factors, among them irregularities in the capstan, in drive pulleys, or in the bearings of the drive motors for the reels or the capstan. Variations in tape tension, or tape stretch caused by temperature and humidity, also can introduce speed errors. Similarly, changes in power supply frequency can affect the synchronous motor that drives the capstan.

Speed variations are minimized by careful design and manufacture of tape transport components and by regulation of recorder power supplies. Speed variations also can be overcome by special components in the more sophisticated instruments. One technique involves a servo-type speed control. In such a system, one recording track is reserved for recording an accurate, constant-frequency signal (usually 60 cycles) along with the data. (Usually this is a 60-cycle tone, which modulates a high-frequency carrier on the control channel. During playback, the control signal is demodulated, and the 60-cycle signal is recovered for the speed control operation.)

During playback, this reference signal is compared with a second reference source of the same frequency in a phase comparator. Any difference in phase produces an error signal that is applied to the drive motor of the capstan to correct the capstan speed so that the played-back reference signal is maintained at the proper frequency.

Most recorders for instrumentation are designed for several applications, involving several different frequency ranges. For this purpose, tape transports normally are made to operate at several different tape speeds. The more common are 1-7/8, 3-3/4, 7-1/2, 15, 30, and 60 inches per second.

Similarly, depending upon the number of tracks to be recorded, the tape transports are designed to handle tapes that are either 1/4, 1/2, or 1 inch wide. Reel sizes vary from 7 to 10-1/2 to 14 or more inches, with tape capacity depending on whether 1-mil or 1-1/2-mil tape is used. A 10-1/2-inch reel will hold 2500 feet of 1-1/2-mil tape, which would yield 16 minutes of recording time at 30 inches per second. The same reel could hold about 3600 feet of 1-mil tape, with about 24 minutes of recording at the same tape speed.

C. Typical Performance Parameters

The operating characteristics which must be considered in selecting a magnetic tape recorder are noted below, and the performance specifications for a typical instrumentation recorder are listed in table VIII.

1. Frequency response. The frequencies that can be recorded and reproduced by the recorder system, usually stated for each tape speed at the normal operating level.

2. Signal-to-noise ratio. Ratio of rms noise level to output signal, usually measured at output of a bandpass filter for the range of frequencies specified above.

3. Harmonic distortion. Harmonic frequencies produced by nonlinear operation of the recorder. The amplitude of the distortion is usually a function of the amplitude of the input signal, and it is expressed as a percentage difference between the input and the recorded signal.

4. Input level. The signal amplitude in volts required to produce normal recording level.

5. Output level. The signal amplitude in volts that the reproduce electronics will deliver into a specified load.

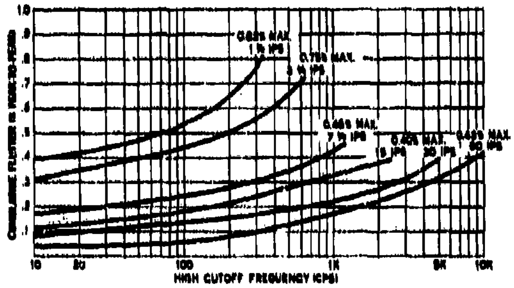
6. Input impedance. The resistive load that the recorder presents to a transducer or other input device.

TABLE VIII. TYPICAL INSTRUMENTATION RECORDER PERFORMANCE SPECIFICATIONS*

Tape Speeds: 60, 30, 15, 7-1/2, 3-3/4, and 1-7/8 ips

Speed Deviation (Long Term): ±0.25% max.

Flutter:



DIRECT RECORD/REPRODUCE MODE

Frequency Response:
RMS S/N Ratio:

Tape Speed (ips)	Bandwidth (+3 db)	S/N Ratio
60	300 cps - 300 kc	32 db
30	150 cps - 150 kc	32 db
15	100 cps - 75 kc	30 db
7-1/2	50 cps - 38 kc	26 db
3-3/4	50 cps - 19 kc	25 db
1-7/8	50 cps - 10 kc	25 db

Input Level:	1.0 volt rms nominal.
Input Impedance:	20K ohms resistive, minimum.
Output Level:	1.0 volt rms nominal, across 600 ohms or greater load.
Output Impedance:	Less than 50 ohms.
Harmonic Distortion:	Less than 1.0% total of a 60-kc signal, at 60 ips.

FM RECORD/REPRODUCE MODE

Frequency Response:
RMS S/N Ratio:
Harmonic Distortion:

Tape Speed (ips)	Frequency Response (+1.0 db)	S/N Ratio	Total Harmonic Distortion
60	0 - 20,000 cps	44 db	1.5%
30	0 - 10,000 cps	44 db	1.5%
15	0 - 5,000 cps	42 db	1.5%
7-1/2	0 - 2,500 cps	42 db	1.5%
3-3/4	0 - 1,250 cps	40 db	2.0%
1-7/8	0 - 625 cps	40 db	2.0%

Input Level:	1.0 volt rms for ±40% deviation.
Input Impedance:	20K ohms resistive, minimum.
Output Level:	1.0 volt rms, across 10K ohms or greater load.
Output Impedance:	1000 ohms, unbalanced to ground.

*Ampex Model FR-100C, Ampex Corp., Redwood City, Calif.

7. Output impedance. The impedance which the recorder will present across the terminals of the following stage.

8. Drift. In FM recording, the percentage of change in carrier deviation, for a specified operating period.

9. Flutter. The variation in tape speed, usually specified as the percentage of cumulative peak-to-peak flutter components actually recorded over the bandwidth at each tape speed.

TIMING DEVICES

Timing devices perform four basic functions in physiological monitoring systems: (1) they supply an accurate reference time base for experimentation, (2) they provide correction signals to correct tape recorder speeds, (3) they produce clock pulses and sync signals for the operation of time-division-multiplexed telemetry systems, and (4) they provide sync signals for separate recorders. The signals produced by the timing devices may be either relative to one another or an indication of absolute time.

1. Reference Signals

All physiological parameters vary as functions of time. To know precisely when the variables being monitored undergo changes, some method of supplying a time base for experiments is necessary. When chart-type recorders are used, the transverse rulings on the chart paper when related to the known constant speed of the chart paper provide the required time base. An alternate method that generally provides greater accuracy makes use of some type of marker, such as an additional marking pen on a chart recorder. In this method, a pulse is generated at a known frequency, generally 1 cycle per second with low chart speeds and 60 cycles per second with high chart speeds.

One source for 1-cps marker pulses is a small synchronous motor with a speed of 60 rpm. The shaft of the motor is connected to a cam or an arm which closes a set of contacts once each revolution, thus providing the desired pulse. This pulse is fed into the marker channel. A highly accurate 1-cps pulse may be obtained by using the 60-cps power line frequency. In one method, the 60-cps signal is fed to a frequency divider, using synchronized, free-running multivibrators. The frequency divider divides the signal frequency successively by division ratios of 3, 4, and 5. A counting circuit, however, such as shown in figure 79, is more reliable. This circuit is a digital counter designed to count to 60. The input pulses are derived by clipping, differentiating, and rectifying the 60-cps power line frequency. The pulses are fed into a binary counter consisting of a chain of six flip-flop circuits. (As is evident from figure 79, to produce a 1-pulse-per-second output, 64 rather than 60 pulses per second must be fed into the input. The additional four pulses are taken from the output of the fourth flip-flop, delayed for approximately 0.005 second by a monostable multivibrator, and combined

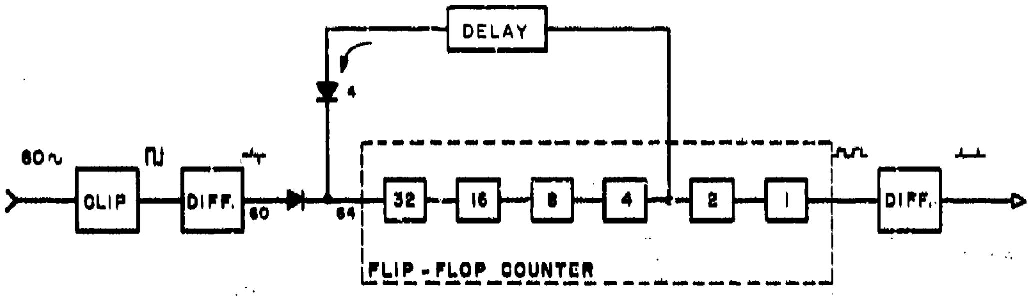


Figure 79. Counting Circuit for 1-CPS Marker Pulses

with the 60-pulse-per-second input.) The final flip-flop is in each state once each second. To provide a short pulse, the output is differentiated and rectified. The accuracy of such a circuit is as accurate as the 60-cps power line frequency, which is on the order of 0.04 percent.*

A 60-cps marker may be taken directly from the power line. The pure 60-cycle sine wave may be used, but a clipped sine wave, which gives approximately square pulses, is a more readable method. In addition to being clipped, the sine wave can be differentiated to give shorter pulses.

II. Correction Signals

Timing signals are used to regulate the tape speed in magnetic tape recording. In order to overcome wow, a 60-cycle sine wave may be recorded on one channel, and by means of phase detection, an error signal produced to correct the motor speed. This is discussed in more detail under Magnetic Tape Recording.

III. Time-Division Multiplex Control

Some type of clock pulse generator must be employed in time-division-multiplexed telemetry systems to control sampling rates and to synchronize decoders with multiplexers.

When electromechanical commutation is used, the sampling is obtained by a motor-driven contactor, and the sampling rate is determined by the speed of the motor. (This is discussed in section V.) When electronic commutation is employed, sampling rates are controlled by electronic pulse generators called clock oscillators. One such

*For power taken from a large, regional grid. Power from a single large generating station is normally frequency stable within a few tenths of one percent, but can be much less accurate during large variations in load.

pulse generator is the free-running multivibrator, an astable device. In this unit the rate of repetition is determined by the time constant of the feedback network. Another device used for the generation of timing pulses is the unijunction transistor oscillator. Schematic diagrams of these devices are shown in figure 80.

If an extremely accurate timing signal is needed, it may be derived from a crystal-controlled oscillator. Crystal-controlled oscillators are inherently more stable than self-controlled oscillators. Accuracy may be improved even more by enclosing the crystal within a temperature-controlled oven. Accuracies as great as 10 parts per million may be expected with such an oscillator. The output of the oscillator is a sine wave; it is changed to short positive and negative pulses by clipping and differentiation or by the use of an amplitude comparator circuit. The positive or negative pulses may then be eliminated by diode rectification.

IV. Sync Signals

If more than one recorder is used, some means must be provided for comparing times on the various records. One method is to have an externally controlled marker channel on each of the recorders, and to have them coupled together so that the records on all channels are marked simultaneously. These time marks may be controlled manually or by one of the time-mark generators described in paragraph I, above.

When data are recorded at stations widely scattered geographically, this method may be impractical. In such cases, a convenient and extremely accurate synchronization signal may be obtained by monitoring the National Bureau of Standards radio

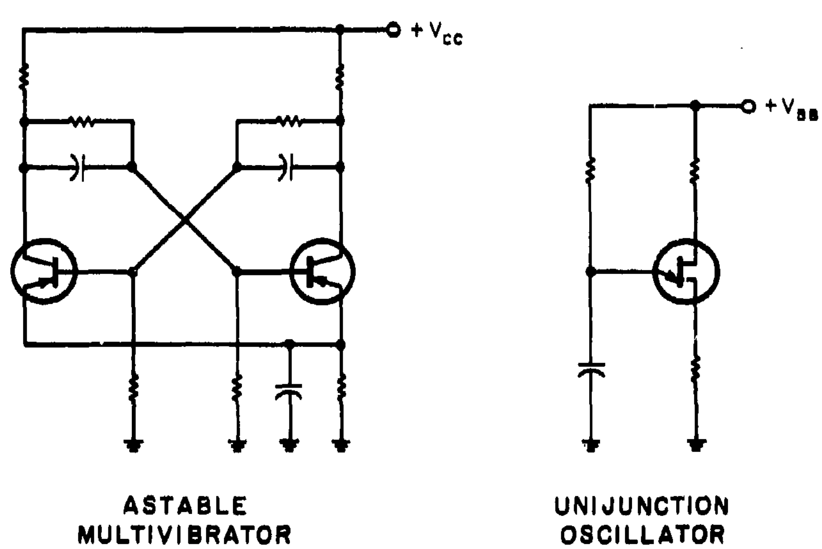


Figure 80. Astable Timing Pulse Generators

stations, WWV or WWVH, or the Canadian Dominion Observatory Station CHU. These stations not only provide very accurate time signals (one per second), but also provide audio signals and accurate carrier frequencies. WWV and WWVH broadcast on frequencies of 2.5, 5, 10, 20 and 25 megacycles; CHU broadcasts at 7.335 megacycles.

POWER SUPPLIES

1. Description

Whether remotely located, such as in aerospace environments, or installed at ground-based facilities, most equipment used in a physiological monitoring system requires some source of electrical power. In some applications a single power supply provides power to all equipments, while other situations require an individual power supply self-contained in each equipment. The general tendency is to use a centralized supply when weight and size requirements are stringent and to use individual supplies in areas when equipment often is used separate from the system.

If the electrical system is land-based, commercial power from large electrical generating stations, burning coal or oil, can be used. The commercial power, normally 115 volts ac at 60 cycles per second, cannot be used directly to power most monitoring equipment. Instead it is used as an input to power supplies in the monitoring system, where it is converted to proper voltage and regulation levels. Aerospace environments demand lightweight, compact sources, and chemical, light, and nuclear energy-conversion devices are used.

All power supplies are basically conversion devices. Batteries convert stored chemical energy into flowing electrical energy; other types of power supplies convert available electrical energy from one form to another, as ac to dc or high voltage to low voltage. When choosing a power supply, the following specifications describing its pertinent characteristics should be considered carefully:

1. Input requirements. The voltage, current, and frequency (if ac) required to operate the power supply at its maximum capability. For a-c sources, the power factor or actual power in watts may also be stated.

2. Output voltage. The levels and type (ac or dc) of voltage delivered by the supply. The a-c waveshape is also specified.

3. Regulation of the power supply (a) with variations of the input voltage, (b) with variations of the steady-state load, and (c) during sudden loads. This latter item is identified as transient response and is stated in terms of recovery time or as capacitance across a load.

II. Electrical Input Power Supplies

As determined by the electrical characteristics of its source and load, an electrical input power supply may be classified according to which of four basic types of energy conversion it effects. The first two types listed below are the most commonly used and will be discussed in more detail:

1. A-c to d-c supplies. This type of supply changes an a-c input signal into a d-c voltage output for sue with amplifiers, electrodes, telemetry systems, etc. These power supplies usually consist of a transformer-rectifier assembly, a filter assembly, and, if required by the power supply specifications, a regulator assembly.

2. D-c to a-c supplies. This type of supply provides a-c power, often needed for transducer excitation, from a d-c input such as a battery.

3. D-c to d-c supplies. This class of power supply changes the input d-c voltage to either a greater or smaller d-c level, with increased stability through regulator circuitry. The input d-c supply may be a d-c bus or a battery supply.

4. A-c to a-c supplies. This power supply may be simply a transformer used to change the amplitude of the a-c input voltage, or it may involve both amplitude and frequency changes, requiring more complex equipment.

A. A-C to D-C Supplies

Electronic devices, whether vacuum-tube or solid-state, require direct current in a variety of voltage ranges for their operation. The most common source of power available is commercial ac at 115 and 225 volts, alternating at 60 cycles per second. This commercial power is converted into the required direct currents and voltages by an a-c to d-c power supply, consisting of a transformer-rectifier-filter assembly.

Figure 81 illustrates a typical power supply of this type. The 115-volt a-c input is stepped up in the transformer, T1, to an alternating current of higher voltage. The blocking action of the rectifier circuit, D1 and D2, converts the a-c to pulsating d-c, and the LC filter, L1 and C1, removes the pulsations from the d-c, producing an essentially constant d-c voltage.

1. Transformers

A transformer consists of two or more induction coils wound about a common core of soft iron. By mutual inductance, a changing current through one of the coils (primary winding) induces a voltage in the other coils (secondary winding). The magnitude of the voltage so induced depends upon the ratio of the number of turns in

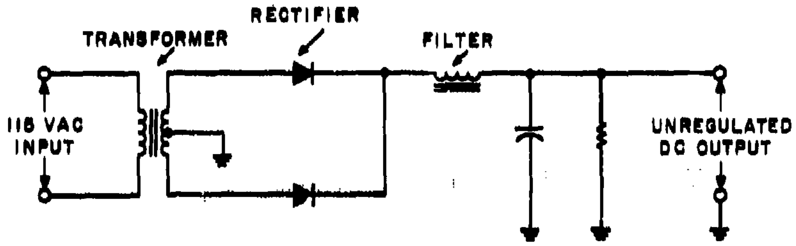


Figure 81. Typical A-C to D-C Power Supply

the primary and secondary windings. There are both step-up and step-down transformers. A step-up transformer provides higher voltages than is supplied to its input; its secondary winding, feeding the rectifiers, has many more turns than the primary winding to which the a-c was applied. Step-down transformers, with relatively fewer turns on their secondary windings, also are used in power supplies to provide lower voltages from the available source for such functions as biasing, tube filament excitation, and lamp circuits.

2. Rectifier Circuits

Rectifiers are electronic devices which conduct or pass current of single polarity. Thus, an alternating current with both positive and negative alterations passes through a rectifier only on alternate half cycles. Whether current flows during the positive or negative half cycle depends on how the rectifier is connected in the circuit.

Most rectifiers used in power supplies are of two types: high-vacuum or solid-state. Both are diode types, consisting of an anode and a cathode, and conduct only when the anode is positive with respect to the cathode. There is a breakdown point at which rectifiers conduct when a negative voltage on the anode is sufficiently high. This is the peak inverse voltage rating of the device, and it should be about three times the rms value of the voltage applied to the device.

Rectifiers also are rated by the d-c load current. Vacuum-tube types can handle up to 275 milliamperes (low power) at 400 or 500 volts dc, or up to 500 milliamperes (high power) at about 2000 volts dc. Solid-state types are rated even higher: germanium rectifiers are rated up to 400 milliamperes, selenium types run to 1000 milliamperes or more, and silicon units range up to several amperes. The current-handling capacity of a rectifier can be increased beyond its rating by connecting rectifiers in parallel.

Rectifier circuits employing a single rectifier are called half-wave rectifier devices, since they conduct the alternating signal from the transformer only during the half cycle that makes the anode positive with respect to the cathode. The

shape of the pulsating output is indicated on figure 82, showing vacuum and semiconductor rectifier configurations. The pulses in the output (the ripple frequency) correspond to the frequency of the ac, and it requires considerable filtering to provide a smooth dc of constant voltage. The half-wave rectifier may be used in some circuits where a small amount of ripple in the voltage can be tolerated; examples are some differential and push-pull amplifiers, switching circuits, and CRT high-voltage supplies.

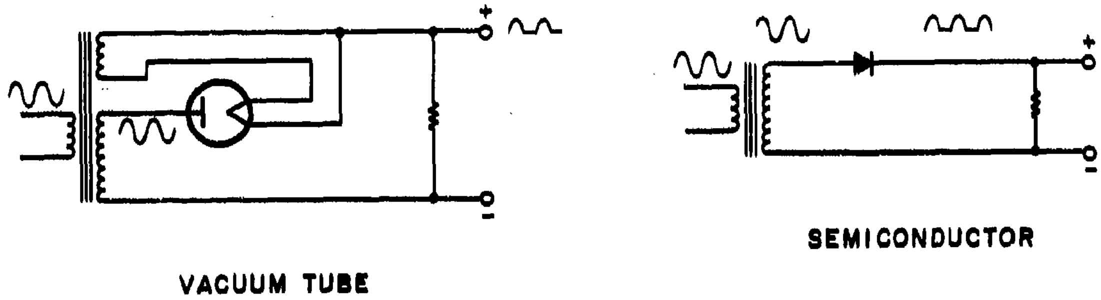


Figure 82. Common Half-Wave Rectifier Circuits

Far more widely used is the full-wave rectifier circuit shown in figure 83. In this configuration two rectifiers are used with a center-tapped transformer. With the center of the secondary winding at ground, the voltages induced between ground and one end of the winding are 180° out of phase with the voltage induced between ground and the other end. Therefore, conduction in the two rectifiers is alternating, and the two half-wave outputs can be combined to produce an output with

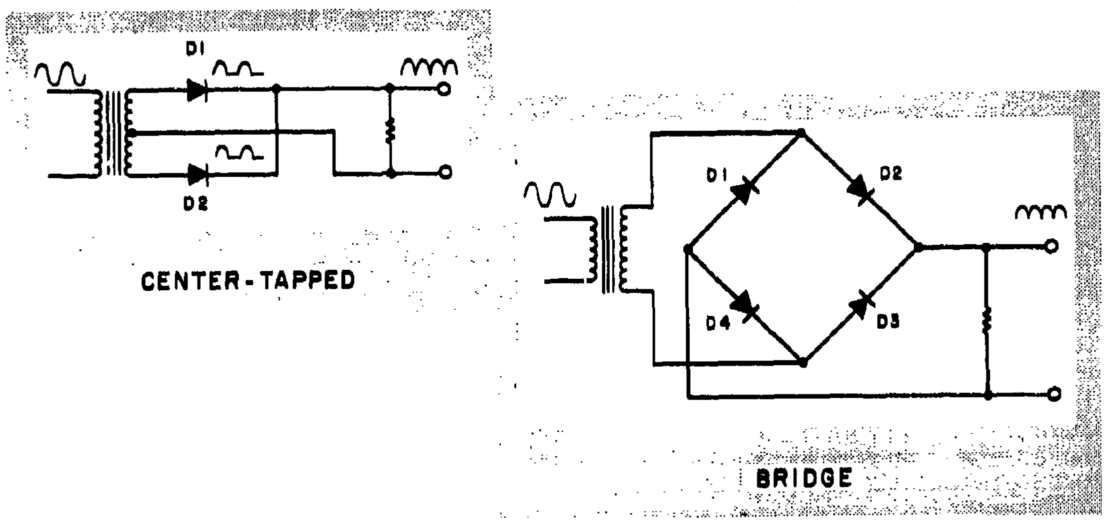


Figure 83. Full-Wave Rectifiers

continuous pulsations. The ripple frequency of the full-wave rectifier is twice that of the frequency of the ac, and it can be smoothed or filtered more readily than the half-wave output to produce a d-c voltage of constant amplitude.

Another type of full-wave rectifier circuit is shown in figure 83. Called a bridge rectifier, it uses four rectifiers and does not require a center-tapped transformer. The diodes are connected so that, on one half cycle of the alternating current, current flows through rectifiers D1, R1, and D3; and on the other half cycle, current flows through D2, R1, and D4. In both instances, current flows through R1 in the same direction, and the output waveform is identical with that of the center-tapped rectifier. The bridge rectifier delivers twice the average voltage of the center-tapped rectifier for the same secondary voltage, but only approximately half the d-c load current.

3. Filters

The output of a rectifier circuit may be thought of as a direct current with a-c components. Filters consisting of capacitors and inductors are used to remove the a-c components, leaving a smooth dc of constant amplitude. The two chief filter configurations are capacitive-input filters (figure 84) and choke-input filters (figure 85), named for the filter element which first follows the rectifier.

Capacitive-input filters have higher output voltages with respect to transformer voltage than do choke-input filters, but the latter type gives better voltage regulation. Voltage regulation is the change in output voltage with changes in load; power supply voltages can decrease with increasing loads (increased current drain) and

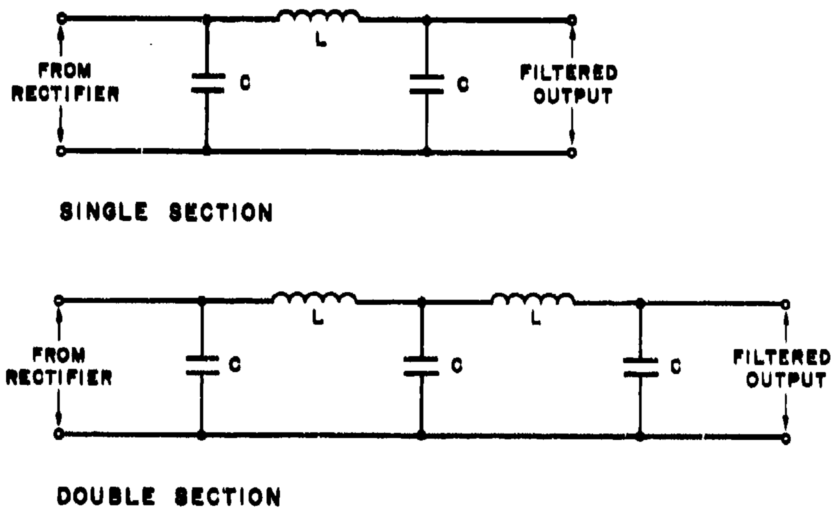


Figure 84. Capacitive Input Filters

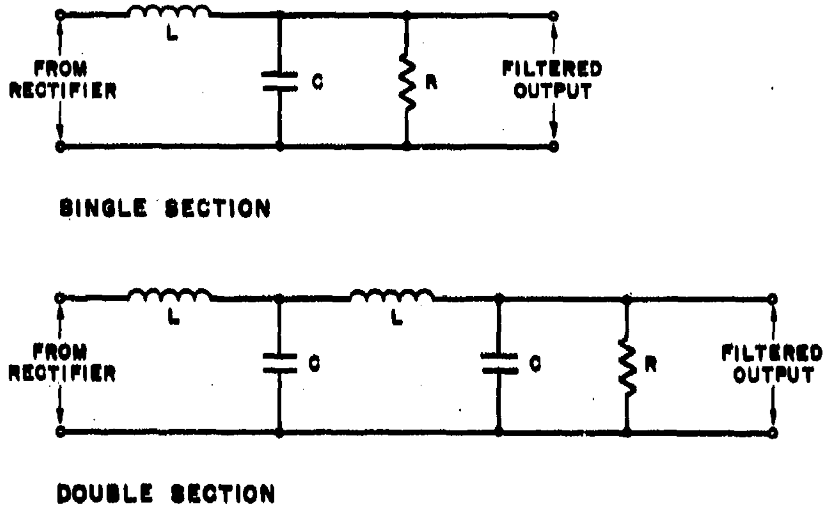


Figure 85. Choke Input Filters

can be excessively high when the load is removed. Some measure of voltage regulation is afforded by the use of bleeder resistors, placed across the output of the power supply. The bleeder provides a minimum constant drain on the power supply to prevent excessive voltages (which can short-circuit or burn out rectifiers, capacitors, and chokes).

4. Voltage Regulators

Each power supply possesses a finite internal impedance. Therefore, its output voltage tends to decrease with increased current drain resulting from changes in load. Also, power supply output voltages vary with fluctuations of the input power source. These variations must be controlled because many types of physiological monitoring equipment require very stable operating voltages.

Power supply output voltages are regulated basically by placing a variable impedance in series with the load so as to form a simple voltage divider. All the current from the power supply passes through the variable resistor and the load, and if the supply voltage rises, there is a proportional rise in the variable resistor and the load. By increasing (or decreasing) the resistance in proportion to the increase (or decrease) in supply voltage, the voltage drop across the load can be held constant.

The simplest method for changing the variable impedance as the supply voltage changes is to shunt the load with a device that will maintain a constant voltage drop over a wide range of current. Two such devices are the gas-filled voltage regulator tube (VR) and the zener diode. Both are shown in figure 86.

The VR tube, a glow-discharge tube, is a diode filled with ionizing gas.

The internal impedance of the tube is inversely proportional to the current flow, and the voltage drop across the tube is practically constant. VR tubes are available for regulation of voltages of 75, 90, 105, and 150 volts. If higher voltages are used, two or more tubes may be connected in series.

The zener diode operates similarly. It is a special type of silicon junction diode with the property of increasing its conductance for increases in current, so that, like the VR tube, its internal impedance is inversely proportional to current.

With the VR tube or the zener diode shunted across the load and voltage applied to the shunt through dropping resistor R1 (as shown in figure 86), any change in the supply voltage will be realized as a corresponding change in the voltage drop at R1, since the voltage drop across the shunt remains the same.

The current and voltage ranges that can be handled by the shunt-type regulators described above are somewhat limited. For handling outputs at higher ranges, numerous regulating circuits are available that use tube or transistor stages in series with the load as variable impedance devices. Vacuum-tube series regulators have higher voltage-regulating capability, from several hundred to over a thousand volts, at currents of several hundred milliamperes. Transistor series regulators work at low voltages typical for solid-state circuitry (6 volts, 18 volts, 30 volts, etc) but are capable of handling several amperes of current.

Figure 87 shows a transistorized series regulator which provides a regulated output that can be adjusted from -6 volts to +6 volts, at currents as high as 4 amperes. (The auxiliary output, an unregulated 5 volts (nominal) at about 50 milliamperes, is used for biasing within the regulator.) The regulated output is stable within 1 percent for an 8-hour operating period, with an input range of from 7.1 to 50 volts.

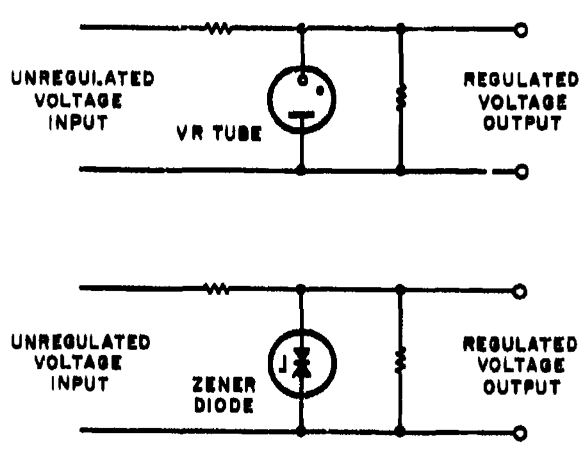


Figure 86. Shunt Regulator Circuits

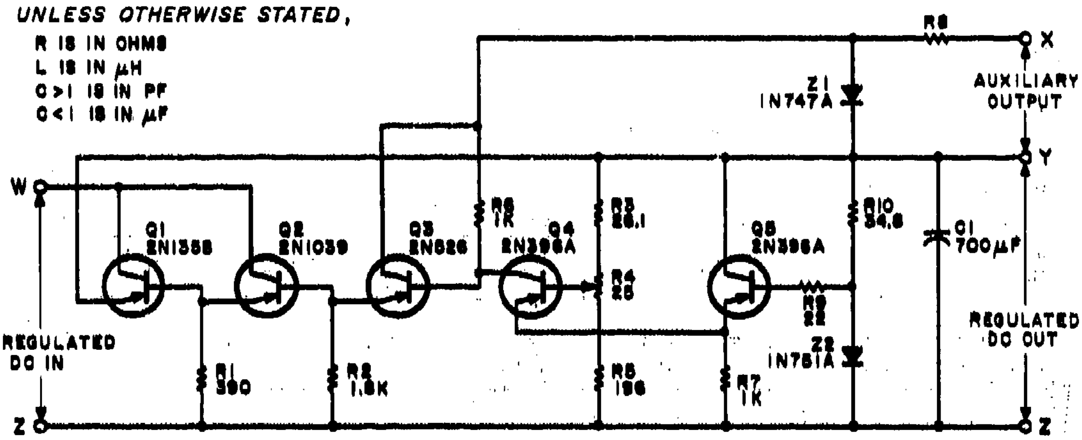


Figure 87. Series Regulator Circuit

Transistor Q1, with emitter followers Q2 and Q3, forms the variable impedance network, and transistors Q4 and Q5 form a differential amplifier which is the comparator or feedback circuit for the regulator. The base of Q5 is tied by diode Z2; the voltage on the base of Q4 is derived from resistor R4. If the load resistance or the supply voltage changes, the voltage on the base of Q4 will change accordingly, and there will be an output across resistor R6, the load for the differential amplifier. This output drives the emitter follower stages, Q1, Q2, and Q3, in the direction necessary to change the input current through Q1, returning the output voltage to the correct level and restoring the balance of the differential amplifier.

B. D-C to A-C Supplies (Inverters)

An inverter is a device that changes a d-c voltage into an a-c output. One type much used in commercial power service is the electromechanical inverter, a rotating machine driven by the d-c source with an a-c generator coupled to the same shaft. This type has also been used in aircraft; in aerospace applications where zero gravity effects prevail, however, the gyroscopic effect of a large rotating converter may present serious problems in navigation and stabilization.

An all-electronic inverter is available for the d-c to a-c conversion in such applications. The d-c source is used to power a multivibrator. The output of the multivibrator is a transformer with a turns ratio that gives the desired a-c output voltage. The frequency of the multivibrator is fixed by the transformer parameters and the supply voltage. (See figure 88.)

Either transistors or silicon-controlled rectifiers may be used to power the multivibrator. In the Project Mercury program three inverters were used, two of which had the inversion capability of supplying 250 volt-ampere outputs at 115 volts, 400 cycles per second (ref. 42).

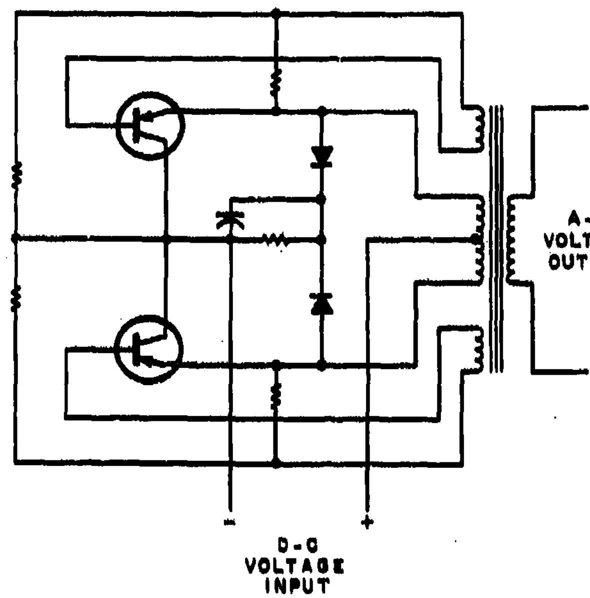


Figure 88. D-C to A-C Inverter

III. Chemical, Light, and Nuclear Input Supplies

In situations where the electrical power developed from large-scale generating stations is not available, namely portable and aerospace applications, the electrical energy must be obtained by some other method. Three conversion methods have been most successful:

1. Chemical-to-electrical, using batteries.

2. Radiant- or thermal-to-electrical, using photoelectric thermoelectric cells.

3. Nuclear-to-electrical or nuclear-to-thermal-to-electrical.

A. Chemical Batteries

The basic configuration of a battery cell consists of two electrodes with an electrolyte between them. A cell with two electrodes submerged in liquid electrolyte is called a wet cell. In dry cells the electrolyte is a paste or is contained in some inert material (special paper, etc).

For both wet and dry cells, the positive electrode contains an oxidizing substance called the depolarizer, and the negative electrode consists of a soluble metal

or metal compound. In the process of discharge, the depolarizer material as well as the negative electrode material and electrolyte are changed through chemical reaction; the depolarizer becomes reduced by reaction with hydrogen, and the negative electrode material is oxidized. The total energy output of a battery cell therefore is a function of the amount of depolarizer material or negative electrode material.

Two classes of batteries are in use: primary batteries and secondary batteries (also called rechargeable or storage batteries). In primary batteries the active material (depolarizer and negative electrode) cannot be regenerated. After discharge, the battery must be replaced with a new one. In rechargeable batteries a charging current (flowing in a direction opposite to the discharge current) restores the depolarizer and the negative electrode through electrolysis. However, the number of charge and discharge cycles is limited and dependent on the design of the battery and the percentage of discharge.

The operating principle of fuel cells is similar to that of primary batteries, except that the depolarizer and the active negative electrode material are supplied continuously from outside the cell. Oxygen or atmospheric air is used as the depolarizer; hydrogen or hydrogen compounds are the active material on the negative electrodes. The electrode itself contains catalytic material to speed up the chemical reactions. The reaction products, water, carbon dioxide, etc, must be removed continuously to keep the fuel cell operating.

The efficient use of batteries requires the consideration of the following factors:

1. Voltage. The voltage of a battery cell depends on the kind of electrode material and electrolyte used. Connecting battery cells in series multiplies the total voltage by the number of cells.

2. Capacity. Capacity is measured in ampere hours or milliampere hours and is the product of the current times the maximum discharge time.

3. Discharge current. Batteries for low or high discharge current are designed differently. Exceeding the specified maximum discharge current causes loss in capacity. The parallel connection of similar cells multiplies the current by the number of cells.

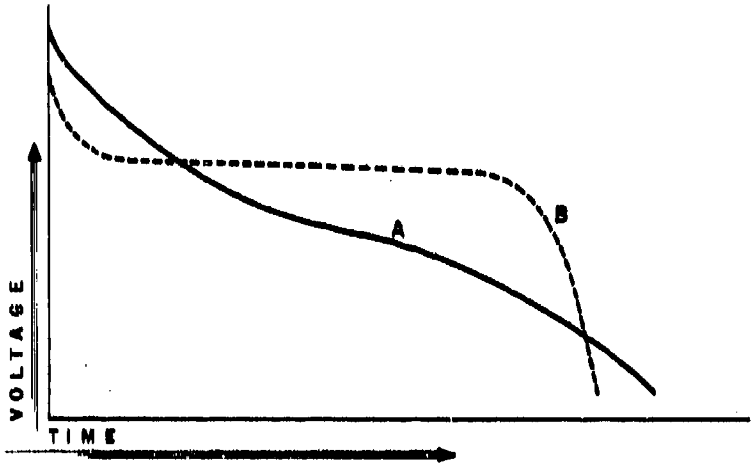
4. Discharge characteristics. The two basic kinds of discharge characteristics available, A and B, are shown in figure 89. The discharge characteristics for different battery types are listed in table IX.

5. Internal resistance causes voltage drops with loading. Internal resistance may be increased with age or through incomplete sealing of the cell. Internal resistance is an inverse function of cell volume, but is also dependent on design factors.

TABLE IX. BATTERY DISCHARGE CHARACTERISTICS

Type of Battery	Electrodes Dopolarizer	Elec- trolyte	Voltage Per Cell	Internal Resistance (Ohms)	Maximum Discharge Current	Energy-to- Weight Ratio	Remarks
Manganese zinc dry cell	Carbon MnO_2 Zinc	NH_4CL	1.5*	0.05–0.3	5–200 ma	Low	Rechargeable
		KOH	1.6*	0.01–0.1	10–500 ma	High	Primary Rechargeable
Mercury dry cell	Carbon HgO Zinc	KOH	1.35**	0.05–0.1	10 ma to 5 amp	Medium	Primary
Silver zinc Silver cadmium	Silver AgO Zn, Cd	KOH	1.6** 1.4	0.001–0.01	Up to 1000 amp	Very high	Rechargeable
Nickel cadmium	Nickel NiO Cadmium	KOH	1.25**	0.01	25 ma to 100 amp	Medium	Rechargeable
Lead acid	Lead PbO_2 Lead	H_2SO4	2.1**	0.001	Up to 500 amp	Low	Rechargeable
Fuel cells	Carbon or nickel oxygen or nickel hydrogen	KOH	0.95**	0.001–0.01	Up to 100 amp	High	Requires oxygen and hydrogen supply

*Per curve A, figure 89. **Per curve B, figure 89.



NOTE:
TYPE B DISCHARGE IS RELATIVELY CONSTANT VOLTAGE.

Figure 89. Battery Discharge Characteristics

6. Energy-to-weight ratio. This ratio is measured in watt hours per pound and depends on the kind of electrodes and electrolyte, as well as the mechanical design of the battery.

B. Light Input Power Sources

The energy contained in radiation in the visible and near-visible wavelengths may be transformed into electrical energy with the use of the solar cell. Actually, this is an indirect transformation of nuclear energy into electrical power, with the light produced by the sun an intermediate step in the conversion process. While the silicon solar cell is extremely reliable and long-lived, its low efficiency, low output, and need for an auxiliary storage device limit its usefulness. The silicon solar cell is about 10 percent efficient in transforming the total incident radiation on its surface into a d-c voltage. Each cell output is only about 0.4 volt, so that a series of them must be connected to provide a usable source of power. For sufficient current output, combinations of the clusters of series-connected cells are connected in parallel.

Because the cells produce an output only when they are being exposed to radiated energy, some means are necessary to provide continuous power in case of momentary loss of sunlight. Often a rechargeable dry cell is used with the solar cell; the dry cell charges during the period the solar cell is producing an output and delivers power to the load when the solar cell is not producing, maintaining a continuous output. In addition, cells connected in parallel must be isolated by means of diodes in the interconnecting wiring. If this were not done, the producing cells would deliver their power to the darkened cells instead of to the load and to the rechargeable

battery during a period when only a portion of the cells were illuminated.

C. Nuclear Power Sources

For long space flights nuclear power sources appear very attractive because they require a minimum of maintenance and most important, the potential energy per pound of fuel is immense. Nuclear energy is changed into electric energy by direct and indirect-conversion methods.

1. Direct Conversion

One type of device, the strontium 90 battery, collects beta particles emitted from the radioactive strontium on a metal electrode, thereby developing a potential between the collecting electrode and the emitting electrode that surrounds the strontium source. Very low currents at high voltages are characteristic of its output.

Another nuclear cell produces a voltage output by ionizing a gas separating the two electrodes of the battery. The gas is ionized by a source of tritium, a hydrogen isotope. The voltage and current characteristics of this nuclear cell depend on the activity of the ionizing source and the materials of which the collecting electrodes are made.

A third nuclear power source depends on the potential created across a semiconductor junction when it is irradiated by a nuclear source. In operation, this is similar to the solar cell, except that instead of the radiation being in the visible light spectrum, nuclear radiation (usually obtained from strontium 90) is the source. The semiconductor junction is ultimately damaged by the action of the radiation upon it, limiting its operation to a period of weeks.

2. Indirect Conversion

In one type of indirect conversion nuclear battery, a phosphor or plastic is bombarded with radiation, and the light generated from this action in turn generates the electrical power across the semiconductor junction. This secondary radiation is not damaging to the junction, providing a lifetime of many years for this cell.

The thermoelectric effect may also be used to generate electric power from a nuclear source. The radiation from a polonium source heats the hot junction of a thermopile. This heat is dissipated by a radiator which forms the cold junction. A unit with as high an efficiency as 10 percent conversion has been produced using this principle.

Section V

DATA TRANSMISSION EQUIPMENT

TRANSMISSION REQUIREMENTS FOR PHYSIOLOGICAL MONITORING

I. Data Transmission and Telemetry - Definition

This section describes the telemetry equipment required for the transmission of quantitative information in a physiological monitoring system. The concept of telemetry, as defined by Webster and as understood by certain branches of engineering, encompasses all functions in a monitoring system: from measurement, through transmission, to ultimate display or recording. This section is concerned with telemetry only to the extent that it furnishes the link between the sensing devices (and their associated modifiers) and the presentation devices of a physiological monitoring system.

II. Elements of the Measurement Situation

In a given measurement situation, there are many interrelated elements that must be considered when determining the requirements for transmission equipment. The most important are:

a. The nature of the information to be transmitted.
b. The type of transmission facilities available.
c. The conditions of transmission.
d. The ideal method for transmitting the information.

The nature of the information generally determines the bandwidth required to transmit the intelligence. The type of transmission facilities available, whether wire or radio, limits the frequency range that can be accommodated. The conditions of transmission involve the number of measurements that must be made (or the number of channels required), the range of transmission, and the kinds of interference that may be encountered. The accessibility of the subject is also a factor, as is the degree and kind of instrumentation he can tolerate. Finally, all of these factors relate to the formation of the transmission signal: can it be sent in its original form or must it be encoded? If encoded, should it be in the magnitude domain or in the time domain?

Modern communications engineering has developed several techniques of encoding, modulation, and multiplexing that are suitable for the needs of physiological monitoring, considering all the above factors. These techniques are reviewed in paragraph III below.

Behind the empiric developments of the communications engineer lies a body of theory setting the limits of the capability of a transmission system. These are the tenets of information theory, or communication theory, and they are discussed briefly in relation to the present context in paragraph IV below.

The remainder of this section, under SYSTEM CONSIDERATIONS, is devoted to a fuller discussion of the types of data-transmission equipment that can be employed in physiological monitoring. For a review of the basic electronics involved in communication and data transmission, see references 3 and 40. Excellent surveys of data transmission and telemetry systems, with more extended treatment than required for this handbook, are contained in references 5, 17, 32, and 51.

III. Basic Transmission Techniques

In the simplest monitoring situation, the subject is at the monitoring and recording station, and perhaps no more than 10 or 20 feet of intercomponent cables and connectors separate transducers, signal modifiers, and display equipment. This type of linkage presents no major transmission problem, but even here the conditions for good transmission must be maintained: components must be matched with respect to input-output impedance relationships, and connectors must be selected for low signal loss and shielded from sources of noise.

In this linkage, signals are transmitted in their original form, i.e., d-c voltages which vary in amplitude. The signals constitute continuous analog data, and they are said to be transmitted in the magnitude domain.

A. Magnitude Domain Transmission

Transmission in the magnitude domain simply means that the intelligence (quantitative information) to be transmitted is represented directly by the amplitude of the signal being transmitted. For faithful transmission, the transmission medium must be capable of passing all amplitude variations at all frequencies (rate of change of amplitude), including zero (dc) or reference level.

1. Direct-Wire Transmission

In direct-wire links (such as that described above), losses in signal amplitude are caused by line resistance and line leakage in direct proportion to line length. Low frequencies, including the d-c or reference levels, may be lost altogether unless d-c amplifiers are used, and high frequencies are attenuated by shunt capacitance in the line. In general this means that the distance a direct-wire linkage can be used is limited, with suitable amplification, to perhaps half a mile.

2. Amplitude Modulation

Wire transmission can be used over longer distances, up to a few miles, by employing the carrier-frequency technique. In this method a constant-frequency signal, called the carrier, is transmitted over the wire link, and the original information-bearing signal is used to change the amplitude of the carrier signal. This process is called amplitude modulation (AM). (Refer to figure 90.)

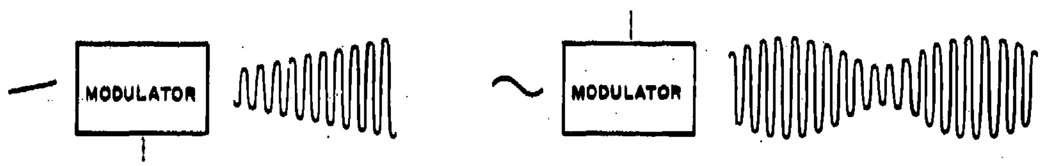


Figure 90. Amplitude Modulation Waveforms

The carrier frequency must be higher than the highest frequency component in the modulating signal by a factor of at least five. The information is now contained in a relatively narrow band of frequencies near the carrier frequency, and it is this band of frequencies that the transmission medium must pass. A-c rather than d-c amplifiers can be employed, and the problems of amplifier drift and base-line stability are no longer important. The principal limitations to long-distance wire transmission using amplitude modulation are increases in variable losses and noise as the line length is increased.

Using amplitude modulation, but with a carrier of much higher frequency, a radio link may be used (rather than a wire link) with greatly increased range. With a carrier frequency of at least several hundred kilocycles, the modulated signal can be fed to an antenna and radiated into space as electromagnetic power. The inherent range capability of such a link is increased greatly, but so long as the modulation remains in the magnitude domain, faithful data transmission is curtailed seriously by system nonlinearities, variable losses, and noise.

3. Limitations of Amplitude Data

The limitations of transmission in the magnitude domain are fairly evident from the preceding discussion. Summarizing, if the information to be transmitted is encoded as variations in magnitude or amplitude, transmission is only as good as the ability of the system to retain those variations faithfully. Simple attenuation, system noise, system nonlinearity, or interference from other signal sources may cause errors in the intelligence transmitted.

a. Attenuation. In a wire system, signal losses are directly proportional to increases in line length. In a radio link, signal losses increase as the square of the

distance. Within limits, attenuation can be offset by increasing the transmitting power. The sensitivity or amplification at the receiving end of the link also can be increased, but this does not improve the signal-to-noise ratio, since any noise present with the information signal realizes the same "improvement."

b. System Noise. All electrical or electronic systems produce noise components which add to the signal passing through the system. These components are discussed later in paragraph IV.

c. System Nonlinearity. Few system components are completely linear. At each amplifier stage, coupling network, etc, the amplitude response is better at some frequencies than others, resulting in distortions in the quantitative information being handled.

d. Interference. Spurious signals all contribute unwanted changes in amplitude-coded signals. Sources include 60- and 120-cycle power lines, transient signals from electrical machinery and switch contacts, bimetallic junction potentials, reflections due to improper impedance matching of transmission elements, and static electricity present in a radio propagation path.

Various techniques are available for optimizing transmission under these conditions, but rigorous fidelity in the transmission of quantitative data more often requires a method of encoding independent of signal amplitude. This method is called time-domain transmission.

B. Time-Domain Transmission

For transmission in the time domain, the intelligence data to be transmitted are represented by the transmitted signal varying with time rather than with amplitude. The variations may be changes in the instantaneous frequency or phase of a sinusoidal carrier, or they may be changes in the characteristics of a pulse-type carrier signal. The various methods for time-domain encoding are noted below.

1. Frequency Modulation

The simplest form of time-domain encoding is frequency modulation (FM), in which successive cycles of a sinusoidal carrier signal are made to vary in period or frequency in accordance with the variations in amplitude of the information signal (see figure 91). This form of magnitude-domain encoding retains the information in continuous, analog form.

A variation of frequency modulation is phase modulation, in which the modulating signal is used to vary the relative phase of the sinusoidal carrier. Phase modulation is seldom used in its pure form for telemetry because of difficulties in transmitting d-c or base-line reference signals.

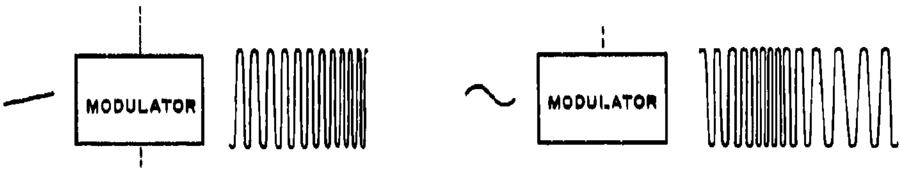


Figure 91. Frequency Modulation Waveforms

2. Pulse Modulation

The preceding types of encoding, both magnitude and time domain, retained the quantitative information in continuous, analog form. The remaining forms of time-domain encoding transmit the information in discrete or discontinuous analog form. The information signal is sampled at periodic intervals and pulses are generated at those intervals which convey, through some characteristic of the pulse, the amplitude of the information signal at the interval. Figure 92 shows a relatively simple sampling technique. If the information signal (A) is passed through a switch that closes only at times T1, T2, T, 2T, 3T, 4T, and 5T, the output of the switch would be a train of pulses (B) whose amplitude correspond to the amplitude of the information signal during the sampling period. Such discrete data may be useful as they are, or, if the sampling rate is high enough, filtering at the receiving end of the link can restore the analog data.

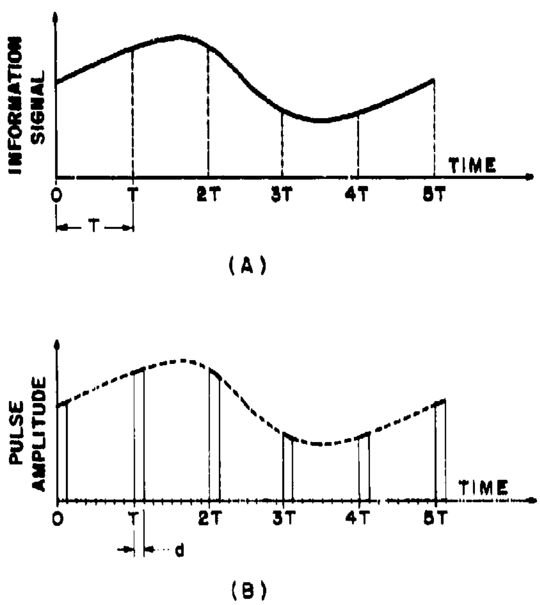


Figure 92. Basic Method of Sampling Analog Data (PAM)

Pulse encoded signals may be sent over wire conductors directly. If the transmission link is radio, the pulse signal is used to modulate the radio-frequency carrier, using either frequency or amplitude modulation (FM or AM). To identify the type of radio transmission, the initials denoting the type of pulse form are combined with the initials of the type of r-f modulation employed; thus, transmission in which a pulse position code is used to frequency modulate the r-f carrier would be identified as PPFM.

a. Pulse-Duration Modulation (PDM)

Pulse-duration modulation is one of two techniques in which the amplitude of the information signal is represented by time. (Pulse-position modulation is the other.) In pulse-duration modulation (PDM), the amplitude at the time of sampling is represented by the duration of the pulse produced during that period, with the pulse duration defined as the time between the rising and falling edge of the pulse. (See figure 93.) This technique is sometimes called pulse-width modulation, or PWM.

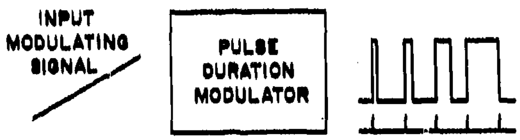


Figure 93. Pulse Duration Modulated Signal

b. Pulse-Position Modulation (PPM)

In pulse-position modulation (PPM), signal amplitude also is represented by time. In this case, the time is the interval between two short pulses generated at the beginning and the ending of the sampling period. (See figure 94.)

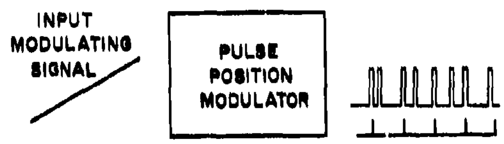


Figure 94. Pulse Position Modulated Signal

c. Pulse-Code Modulation (PCM)

Signal amplitudes also may be represented by code groups. This

technique requires that the information signal first be digitized, and the digital value then be encoded. With PCM, discrete digital rather than analog data are transmitted. Decimal digital data have been transmitted by such dot-dash codes as Morse code. Far more convenient and common in modern telemetry techniques is the mark-space system of binary digital coding. A typical binary digital code is shown in figure 95.

d. Pulse-Amplitude Modulation (PAM)

Pulse-amplitude modulation (PAM) is a special case of pulse coding because it is a magnitude-domain technique: the signal amplitude is represented by the amplitude of the pulse produced during the sampling period (figure 92). However, since this technique entails periodic sampling by relatively simple techniques, it lends itself to use with other time-domain components, and is often used as an intermediate signal form in multiplex systems, with the final transmission being in FM or some other time-domain form.

C. Multiplexing

The encoding techniques described above apply to all data transmissions, whether a single information signal or many are involved. Where information signals from more than one data channel are to be transmitted, it is convenient to combine them so that they can be transmitted over a single transmission channel or link. This process of multichannel telemetry is called multiplexing. Two basic methods of multiplexing are employed, each combining the data channels in such a way that they can be separated again at the receiving end of the telemetry link. These are frequency-division multiplexing and time-division multiplexing.

1. Frequency-Division Multiplexing

In frequency-division multiplexing, several carrier signals (called subcarriers) are amplitude or frequency modulated, each by a separate information signal. Each subcarrier occupies a separate frequency band, usually in the audio range of 400 to 70,000 cycles per second. The modulated subcarriers then are mixed together (added) for subsequent transmission.

In a wire-carrier system, the composite signal may be sent directly. In a radio telemetry system, the composite signal frequency modulates a high-frequency r-f carrier. (If both the subcarriers and carrier are frequency modulated, the system is termed FM/FM; if the subcarriers are amplitude modulated, the system is termed AM/FM.)

At the receiving end, the individual channels are separated by bandpass filters for demodulation. In the radio system, of course, the composite subcarrier signal first must be separated from the carrier by FM detection.

DECIMAL	BINARY CODE	WAVEFORM
0	0000	
1	0001	
2	0010	
3	0011	
4	0100	
5	0101	
6	0110	
7	0111	
8	1000	
9	1001	
10	1010	
11	1011	
12	1100	
13	1101	
14	1110	
15	1111	

Figure 95. Pulse Coding for Binary Digital Data

2. Time-Division Multiplexing

Whereas frequency-division multiplexing is true simultaneous transmission of separate channels (each of which can contain continuous analog data), time-division multiplexing consists of sequential, cyclic transmission of discrete, discontinuous data. This multiplexing technique is best illustrated by the operation of a rotating commutator, in which the several input signals are switched sequentially onto a common output channel. The pulse-modulated, multiplexed output may be transmitted directly

over wire, or used to modulate a high-frequency r-f carrier. At the receiving end, the modulated pulses are separated back into individual channels by a decommutator that is synchronized with the sampling, rotating commutator. Figure 96 illustrates this technique in simplified form.

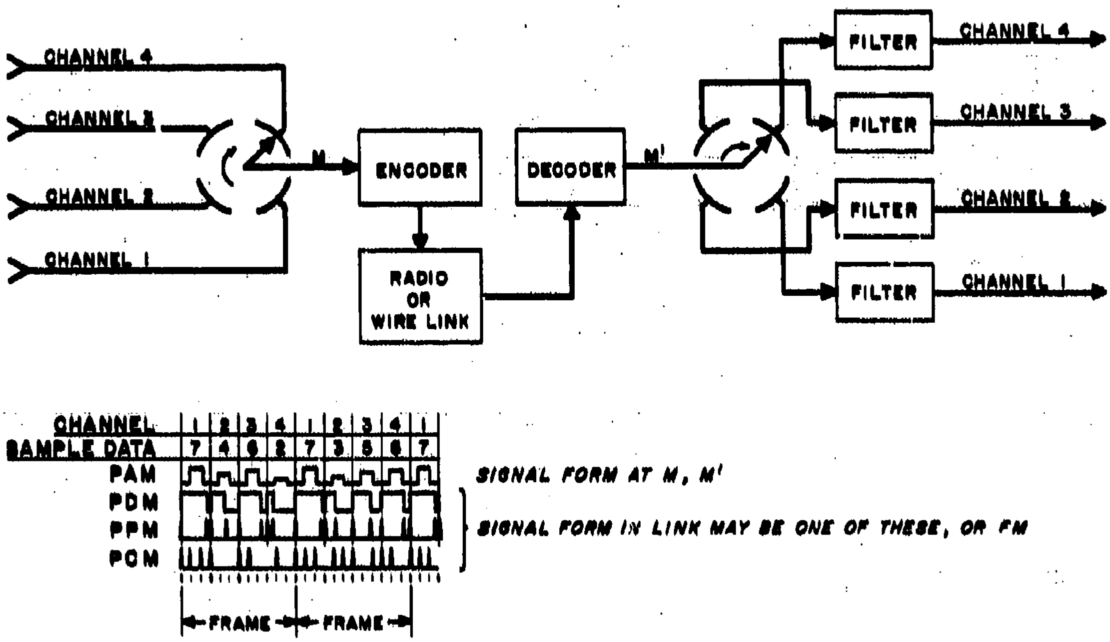


Figure 96. Simplified Scheme for Time-Division Multiplexing

3. Compound Multiplexing

Occasionally, both frequency- and time-division multiplexing are used in the same system. A common technique called subcommutating involves the time-division multiplexing of one or more subcarriers in a basically frequency-multiplexed (FM/FM) system. Thus, several channels of slow-varying variables, such as temperature, might be commutated to frequency modulate a single subcarrier in an FM/FM system, while other channels, requiring continuous data transmission, would each module a separate subcarrier.

IV. Information-Bandwidth Considerations

It has been stated, in paragraph II above, that the bandwidth of a telemetry system is generally determined by the nature of the information being transmitted. Actually, until relatively recent years data transmission or communications engineers did not concern themselves with the information they were asked to transmit, but only the electrical signals which conveyed that information: their waveforms and frequencies,

and the equipment fidelity required for their reproduction.

The deeper investigation into the actual information content came about, in time, with the need for greater communication capacity (more channels, for more communicants) and for better (noise-free, more reliable) communication systems. It was helped along by the investigation of pulsed signal techniques (radar) accompanying World War II.

Early concern for the quantity of information was related to the continued improvement of telegraphic communication. Following the work of Nyquist and others, R.L.V. Hartley in 1928 developed an expression for the quantity of information as a product of bandwidth and time. He stated that a message of M symbols, each chosen from an "alphabet" of N symbols, has N^M possible sequences. Hartley then said that the quantity of information, H, is best expressed as the logarithm of those possibilities, that is, H is proportional to M log N.

This diction -- symbols, choices, alphabets -- is central to the basic definition of information as a choice of one of a number of possible conditions, or as a message containing a certain number of symbols chosen from a prescribed alphabet of symbols. Thus, one way to describe the message contained on this page is as a sequence of words chosen from a dictionary, or of letters from a 26-letter alphabet. The simplest choice of course is one between two possibilities, such as yes-or-no, on-or-off, 1-or-0. This is the principal behind the system of binary digital coding shown in figure 95, and the binary digit, or bit, is the basic unit of information. By such a system it is possible to describe any number of possibilities as a sequence of such binary or two-possibility choices. Thus, any entry in a quarter-million-word vocabulary or dictionary can be encoded by no more than 18 such bits (2^{18} is greater than 250,000).

In more immediate terms, consider how the above factors relate to the transmission of the time-voltage information contained in an electrical analog signal obtained from a biological transducer. Consider a signal that is band limited to W cycles per second, and that is sampled twice per cycle or once every 1/2W seconds, as shown by the sampling ordinates in figure 97. At any time t, or at any sample ordinate, the signal will be at one of N levels of amplitude. By Hartley's relationship, each ordinate then contributes log N bits of information.

For a time interval T, there will be 2WT sample ordinates, each at one of N levels; thus, there are N^{2WT} possible sequences in interval T. The information rate, H, is the log of the number of possibilities, N^{2WT}, or (for log base 2)

$$H = 2W \log_2 N \text{ bits per second.}$$

The total quantity of information in T seconds is

$$H' = 2WT \log_2 N \text{ bits,}$$

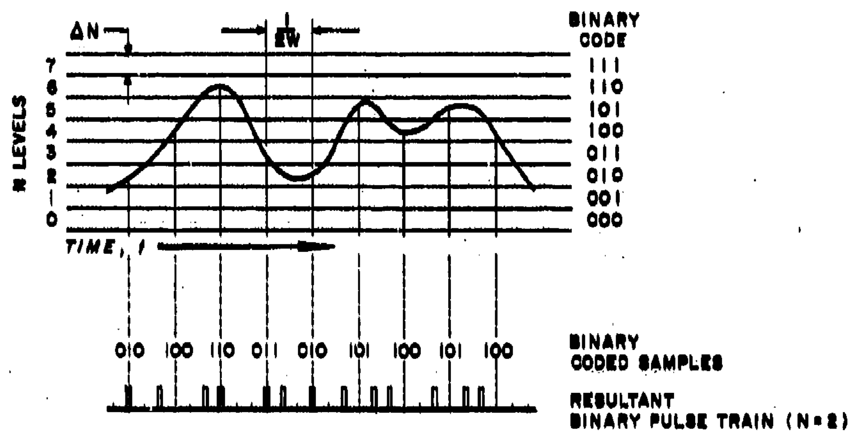


Figure 97. Sampling and Quantization of Continuous Signal

which in turn can be expressed as the information content of each ordinate ($\log_2 N$) times the number of ordinates per second (2W) times the time (T).

Substituting some possible values, if the bandwidth, W, of the signal is 100 cps, and the number of amplitude levels, N, is 8, the information rate, H, becomes

$$
\begin{aligned}
H &= 2W \log_2 N \\
&= 2(100)\ (\log_2 8) \\
&= 600 \text{ bits per second.}
\end{aligned}
$$

Interestingly, if the eight amplitude levels were in turn encoded with a 3-bit binary code, the information rate would remain the same; while the number of sampling ordinates (2W) is effectively trebled, the information content of each ordinate ($\log_2 N$) would be only one-third as great, since N is now reduced from 8 to 2, and $\log_2 N$ from 3 to 1. Thus, a message source requiring relatively low resolution (two levels) but relatively wide bandwidth (600 bits per second) has been exchanged for one requiring high resolution (eight levels) but narrower bandwidth (200 bits per second), with no change in the information rate.

Hartley's treatment of information capacity is the foundation of modern information theory. Its single greatest weakness is failure to account for the effect of noise upon communication. This aspect of the problem has been treated in the statistical theory of communication. Such work relies heavily on the mathematical science of probability; full treatment is beyond the scope of this handbook, but the interested reader is referred to general references (ref. 10, 37) and basic engineering texts (ref. 17, 45) for further discussion.

A fundamental law of information theory, expressed by Claude Shannon, defines

the theoretical limit of the capacity of a communications channel, relating bandwidth and signal-to-noise ratio, as follows:

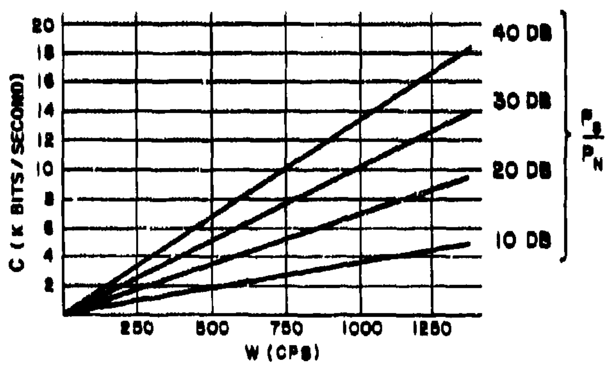
$$C = W \log_2 \left(1 + \frac{P_S}{P_N}\right)$$

where C is the capacity in bits per second,
P_S is the signal power, and
P_N is the power of white (Gaussian or randomly distributed) noise.

This relationship is shown graphically in figure 98, for several different S/N values. The direct relationship between capacity, bandwidth, and power, is apparent.

Thus, doubling channel bandwidth will double capacity directly. More significantly, for a constant capacity, doubling channel bandwidth will permit a much greater reduction in power required, since it is the log of the S/N expression that is halved.

Conversely, operating on the signal power to effect greater capacity or lower bandwidth requires proportionally greater changes in power. It remains at present preferable to operate on the bandwidth, or to operate on the information directly (by coding and information compression techniques) to reduce the required capacity.



$$C = W \ \mathrm{LOG}_2 \left[1 + \frac{P_S}{P_N}\right]$$

Figure 98. Information Capacity of a Channel

The noise power figure employed in Shannon's expression is predicated upon randomly distributed (Gaussian) noise -- the kind of disturbing electrical signals present to some degree in all electronic (specifically data transmission) equipment because of random electron movement in conductors. Such noise sets the limit of the resolution of a system, that is, of the number of levels into which a signal can be quantized. Hence,

it ultimately limits the information rate and capacity of a channel.*

A simple illustration is given in figure 99. Part A shows a binary coded signal, with levels N_1 and N_2, and again the same signal with a noise signal of given level imposed upon it. Part B shows the same noise signal imposed upon a five-level signal, where the noise level approaches ΔN; the likelihood of error in reproducing a given level is self-evident.

There are several possible techniques of improving system reliability and resolution in the presence of noise. One technique, where delayed transmission is possible, is to record the information signal on magnetic tape, and then play the tape at a much lower speed into the transmitter. All frequency components of the information signal will be proportionally reduced, so that the S/N ratio for a given bandwidth will be significantly increased. (Or more information can be sent over the same bandwidth channel with no reduction in S/N.)

Conversely, there are systems in which transmission rates are much greater than original information rates (advantageous where abnormally high S/N ratios obtain). And pulse codes of ever-increasing complexity are being devised, not to remove error

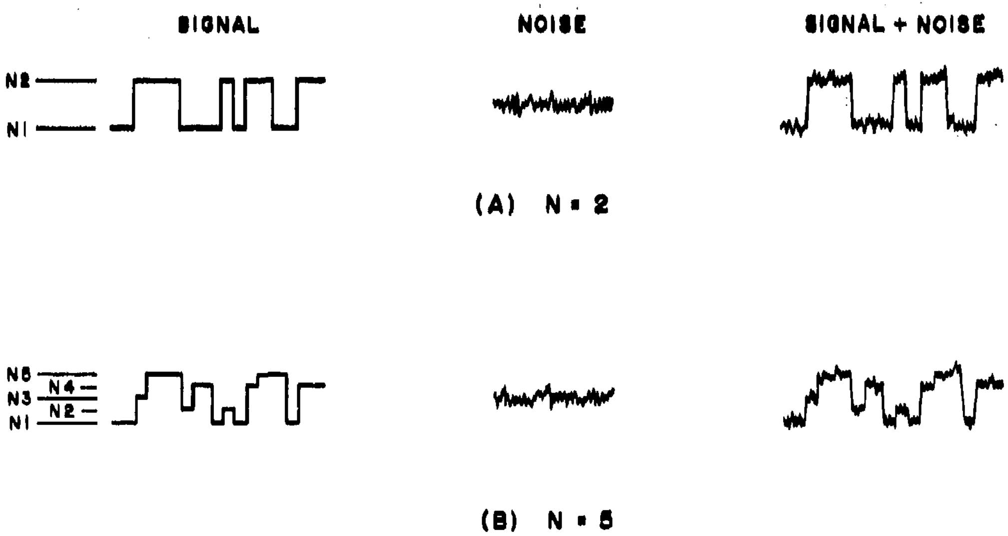


Figure 99. Effect of Noise for Various Values of N

*Practical limits also obtain from non-Gaussian noise that may affect a particular transmission link. Thus, sporadic noise from static discharges, or multichannel crosstalk, or drops in S/N ratio due to signal fading over long distances, must be dealt with independently.

from the transmission, but to detect it so that the original information can still be recovered.

All such techniques entail added equipment complexity, and the cost of the advantages gained must be weighed in terms of money and of equipment size, weight, and power consumption.

SYSTEM CONSIDERATIONS

1. Wire Systems

Data may be transmitted over wire systems in two ways -- either by direct transmission or by carrier transmission. In direct transmission, signals may be sent in their original form or in one of the pulse-encoded formats over relatively short distances. For longer distances, AM or FM carrier transmission is used.

A. Direct-Wire Transmission

Basically, a direct-wire transmission link consists of the necessary length of conductor between system components, as, for example, between a transducer and the preamplifier in the input of a recorder. However, as the length of such a link increases, there are problems in transmitting without loss of intelligence. The simple conductor may have to be supplemented with additional circuit elements and components; a typical direct-wire d-c link is shown in figure 100.

1. Components

The circuit in figure 100 represents a single channel of a multichannel direct-wire link that is used in the medical teaching center at Baylor University (ref. 14). The link spans as much as a quarter of a mile in transmitting experimental measurement data to various display locations in the center. The circuit, because of the distance spanned, includes amplifiers, shielded conductors, and various component-matching and coupling circuits.

a. Amplifiers

D-c amplification is necessary to ensure a good signal-to-noise ratio and adequate power at the receiving end to drive the terminal display or recording equipment. The d-c amplifier used in figure 100 is a power amplifier of the type used to drive the pen motors in galvanometric recorders. It drives the lines at peak signal levels (near 100 volts) into 3000-ohm terminations. The theoretical signal-to-noise ratio is 10,000 to 1.

A preamplifier also is shown that may be thought of as part of the input to the transmission system, since its prime function is to raise the transducer output to the input level of the d-c amplifier.

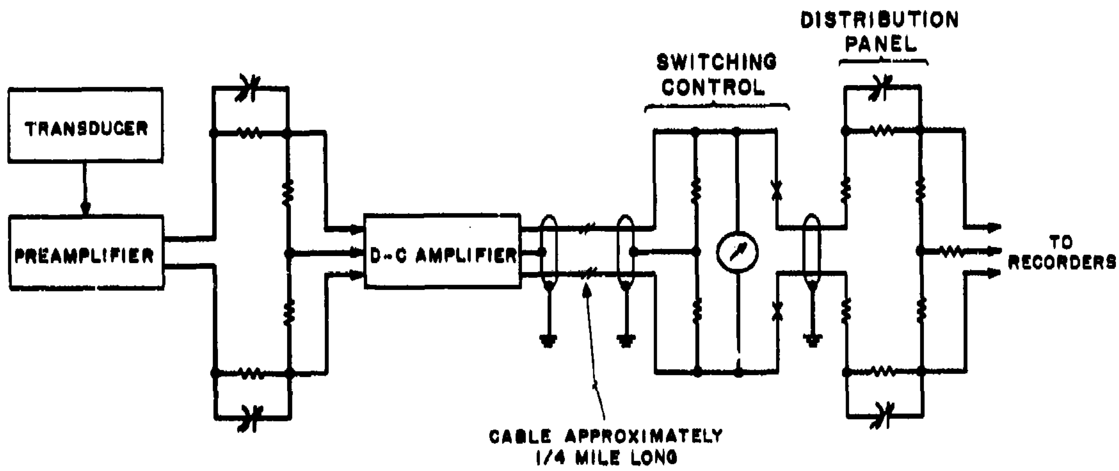


Figure 100. Single Channel of a Direct-Wire Link

b. Attenuators

Having amplified the signal level to improve the signal-to-noise ratio during transmission, the level at the receiving end may have to be reduced with attenuators to avoid overdriving the recording or display equipment. Attenuators are resistive voltage-dividing networks, such as that shown in the distribution panel in figure 100.

c. Filters

Filter circuits are used to match frequency ranges between parts of a circuit. The circuits usually are resistance-capacitance inductance couplings which block frequency components in a signal that are above, below, or between certain predetermined limits. The variable capacitors used in the voltage dividers in figure 100 are filtering elements; they can be used to adjust the frequency response of the link, which ranges from 0 to 5000 cps.

d. Impedance-Matching Devices

Input and output impedances between components must be matched properly for the most effective transfer of signal power in a measurement system. Refer to the discussion of these various components on pages 59 and 71.

2. Selection of Wire and Cable

The transmission of d-c or very-low-frequency a-c signals over direct wire links is limited chiefly by line or loop resistance and leakage resistance, while

the transmission of high frequencies is limited by line capacitance. For the construction of open-wire lines, standard AWG and BWG copper wire with resistances of about 2 to 6 ohms per mile generally is used. When making computations to determine linear line resistance, the entire loop resistance must be considered. For example, if an open-wire line is to be operated between two locations 3 miles apart and the line is constructed of No. 6 AWG copper wire (with resistance of 2.09 ohms per mile), the value of the line resistance (since the current must return to the transmitting point over a closed circuit) is $2 \times 3 \times 2.09 = 12.54$ ohms. The linear resistance of open-wire line, however, frequently is a less important factor than leakage resistance.

Leakage resistance acts as a shunt across the terminals of the telemetering link. This resistance varies with weather conditions, and may affect the accuracy of some instruments materially. The leakage, depending upon the insulation resistance of the wire and its supports, is low in dry weather and high in wet weather. Therefore, open-wire lines should be designed on a basis of wet-weather conditions, when resistance is lower. Convenient values to use as approximations are 0.2 megohm per mile (between line and ground) and 0.4 megohm per mile (between conductors).

Open-wire lines cannot be used where the system is subject to electrical interference from external sources, such as power lines and large electrical equipment. For such applications, shielded cable conductors must be employed. Shielded cable conductors, unlike open-wire lines, present no leakage resistance problem. The insulation resistance of cable conductors in good condition usually is high enough so that, neglecting leakage, the determining factor of the operating range at low frequencies is simply the linear resistance of the cable loop. Some of the commonly used AWG cables have loop resistances as low as 11 ohms per mile (AWG 13) and as high as 200 ohms per mile (AWG 26) (ref. 5).

B. Wire-Carrier Transmission

A wire carrier system requires, in addition to the wire link, a carrier transmitter and receiver, plus modulation and demodulation circuitry. Most of these components are described elsewhere in this section. Actually, most applications requiring modulation and carrier conversion are more suitably implemented with an r-f link.

There are, however, certain wire-carrier techniques that, using existing facilities, can serve a physiological monitoring application without incurring the major expense of line installation.

1. Carrier-Current Systems

Although power lines are designed for 25- to 60-cps operation, they can be used for telemetry links using carrier frequencies ranging from 20 kc to 300 kc, and are capable of handling up to 18 different data channels simultaneously. Line losses can be reduced by suitable design to 0.2 db/1000 feet for 20-kc signals, and to 0.9

db/1000 feet for 300-kc signals (ref. 5).

Wire-type interconnecting circuits are either grounded or ungrounded. The grounded circuits include a one-wire conductor and the earth (ground) as a return path, and a two-wire balanced grounded type. The ungrounded circuit has two distinct wire conductors. In general, two-wire circuits are preferred to grounded circuits; the latter type is subject to ground-potential differences, such as those caused by large electrical systems or magnetic storms. At locations a few hundred miles apart, differences in ground potentials greater than 500 volts may be experienced. Two-wire lines, while not subject to such disturbances, are exposed to the effects of static atmospheric conditions. Both grounded and two-wire lines are subject to severe disturbances when exposed to abnormal operating conditions if located near power-transmission lines (ref. 5), but the grounded system is more adversely affected.

Shielded coaxial cable circuits avoid most of these difficulties. Although relatively expensive, they should be used whenever maximum reliability is required.

C. Installation of Wire Links

A transmission line for physiological telemetering purposes should be installed carefully, since stable lines result in efficient operation with minimum breakdown. All joints and splices should have sound mechanical and solder connections, particularly in those installations operating with low energy levels. Temperature and weather conditions (including wind vibration) also must be considered, since some telemetering systems depend on unvarying line conditions for accuracy.

When ground circuits are used, the system should not be grounded in the vicinity of a high-voltage station ground. If trouble develops in the power system, the disturbances may impair the operation of the telemetering equipment. These disturbances can be minimized by grounding the telemetering system at a point as remote as possible from all power systems.

II. Radio Systems

A. Equipment

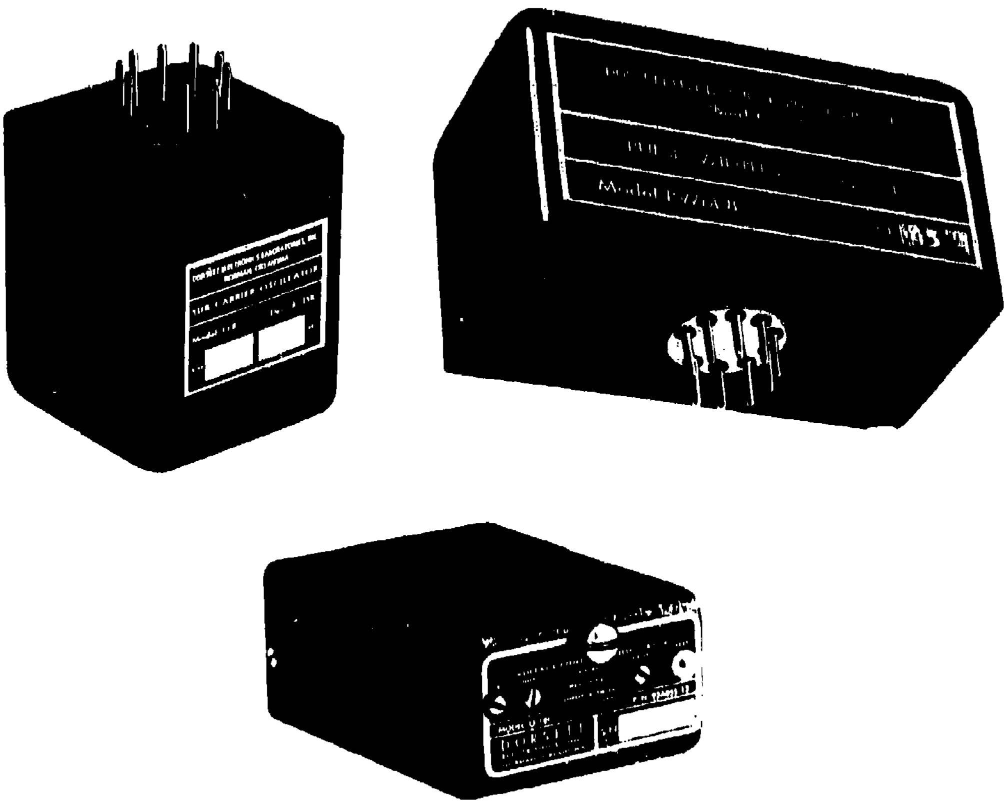
1. Telemetry Transmitters

The telemetry transmitter generates the r-f carrier, modulates the carrier with the intelligence signals, and then amplifies the modulated carrier to provide sufficient power to radiate the r-f signals from an antenna.

Generally, telemetering transmitters should be small, light, reliable, and undisturbed by ambient conditions, and should consume little power. The lightweight and low-power-consumption requirements may not be important in some applications.

Transmitters with direct frequency-modulated oscillators suffer considerably from frequency drift because of temperature variations. This frequency drift usually is small enough so that it can be compensated for by receivers having automatic frequency control features. However, a drifting transmitter frequency wastes space in the restricted r-f spectrum available for telemetering. Therefore, crystal-controlled transmitters with frequency and modulation characteristics that are relatively immune to ambient temperature variations are becoming increasingly popular in telemetering applications.

Ambient vibration can introduce a considerable amount of spurious modulation in a transmitter. This is especially true in FM and PM transmitters, since changes in frequency often result when the distributed parameters of high-frequency circuits are disturbed by vibrations. Air-dielectric capacitors, stray wiring capacitance, air-core inductors, and the interelectrode capacitance of vacuum tubes are a

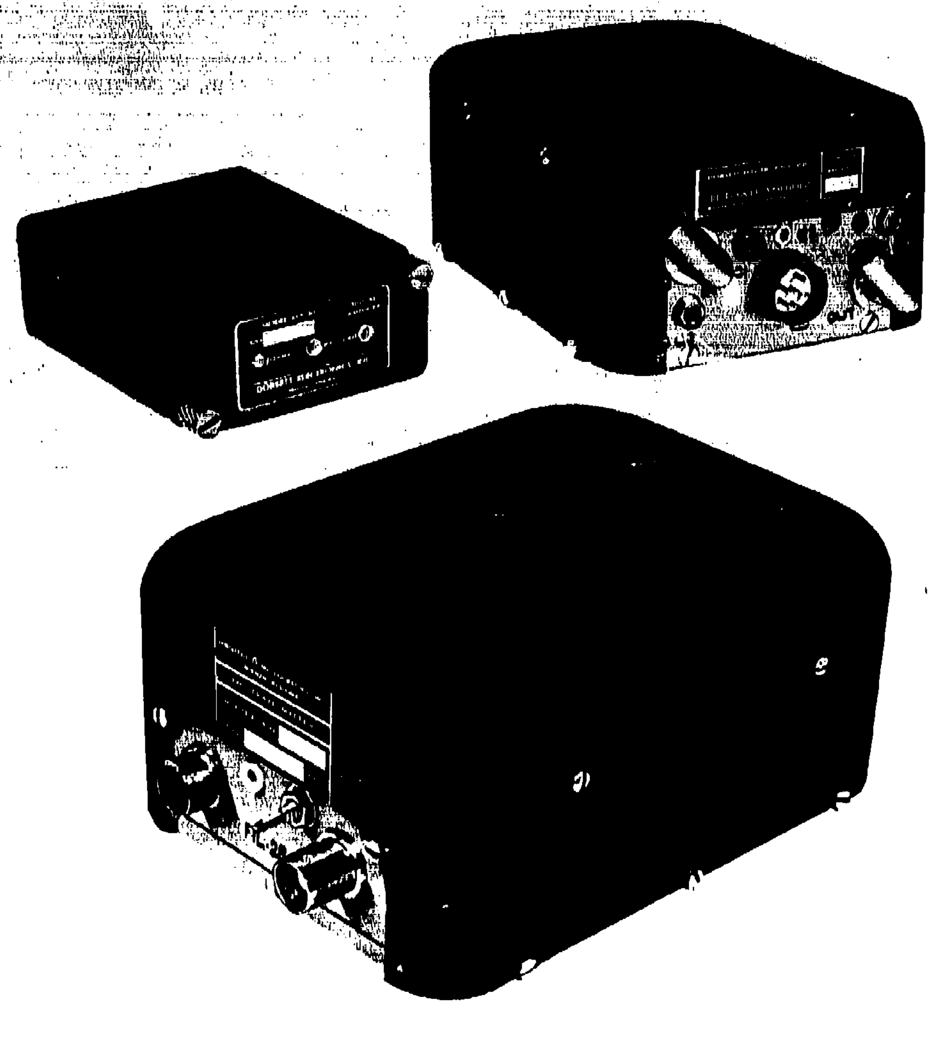


Dorsett Electronics, Inc., Norman, Okla.

Figure 101. Typical FM/FM Telemetry Components (Part I)

few examples of frequency-determining parameters that are susceptible to ambient vibrations.

In standard FM telemetering links employing a maximum frequency deviation of 125 kc, the frequency modulation resulting from ambient vibration usually can be held to a tolerable value if the total frequency multiplication in the transmitter is small. However, when a low-frequency oscillator and a large frequency multiplication is used to produce the desired output frequency, a small amount of vibration-excited



Dorsett Electronics, Inc., Norman, Okla.

Figure 101. Typical FM/FM Telemetry Components (Part II)

frequency modulation in the early stages causes an intolerable amount of spurious modulation in the output. Shock and vibration mountings should be used with such units.

FM radio links are commonly used in r-f telemetry. (They are standard for the 216- to 260-mc telemetering band.) Transmitters with excellent frequency stability (obtained in many cases by crystal-controlled oscillators) are used in order to use the available frequency allocations efficiently. FM telemetry components are available as system building blocks from many manufacturers, so that many physiological applications can be satisfied with off-the-shelf telemetry hardware. Typical units for FM/FM are shown in figure 101, including a subcarrier oscillator, pulse-duration modulator, mixer amplifier, transmitter, and power amplifier.

The power amplifier shown in figure 101 produces 25 watts of r-f power in the 216- to 260-megacycle band, with an input of 2 watts or less from a transmitter. The system designer determines whether power amplifiers such as these are necessary, or whether transmission can be completed with the relatively low-power transmitter outputs available (from several milliwatts to several watts). If the transmitter output is very low (just a few milliwatts) and the power required for transmission is high (near 100 watts), an intermediate power amplifier may be necessary to raise the transmitter output to a level (1 or 2 watts) sufficient to drive the power amplifier.

Typical of the simpler transmitters useful in physiological monitoring is the 3-stage FM/FM unit shown in block diagram form in figure 102. The first stage is a crystal-controlled oscillator, with biasing and modulation controlled by two varicap diodes. The second stage is a third-harmonic amplifier, and the third, output stage is a fourth-harmonic amplifier, so that the oscillator frequency is multiplied by a factor of 12. With modulating frequencies of up to 30 kc, the carrier deviation is 60 kc about a center frequency of 228.2 megacycles.

This transmitter was designed to accept 12 subcarrier channels of low-level physiological data, all of which are applied in parallel to the input mixing resistor. It delivers more than 1 milliwatt power directly to an antenna for a transmission range of about 1 mile. The entire transmitter, with signal-conditioning circuitry and battery power supply, fits into a package 8 x 3-1/2 x 1/2 inches.

2. Telemetry Receivers

The telemetry receiver amplifies the r-f signal intercepted by its antenna, converts it to a lower frequency, and then removes the r-f carrier. The remaining signal is identical to that used to modulate the carrier generated by the r-f transmitter. If multiplexing was used, the modulating signal is then separated into its original components to reproduce the separate channels of data.

If time-division multiplexing was employed, the output signal is applied

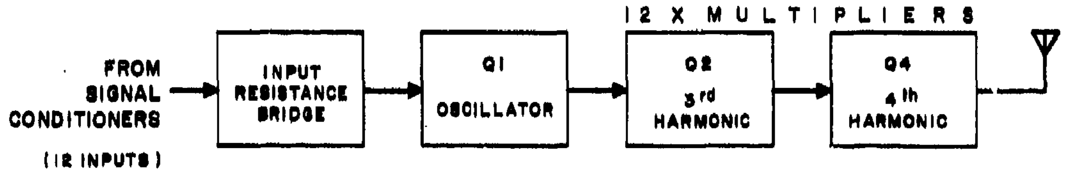


Figure 102. Simple Three-Stage FM Transmitter

through a bank of filters for separation of the individual subcarriers, and each subcarrier then is applied to a discriminator for removal of the subcarrier frequency, yielding as outputs (1) the data signals originally applied to an FM/FM modulating scheme, or (2) pulse trains such as are derived from pulse-type subcommutation. The latter signals must of course be applied to appropriate decommutation circuits for recovery of the data so encoded.

The components of a typical FM/FM receiver are shown in figure 103. The r-f amplifier is a high-gain circuit which raises the level of the signal from the receiving antenna. Its output is mixed in the converter with the output from the local oscillator. The local oscillator is frequency controlled, at a frequency which is either above or below the frequency of the r-f signal being received. The output of the converter is at a lower, intermediate frequency; it is the difference between the r-f and local oscillator frequencies.

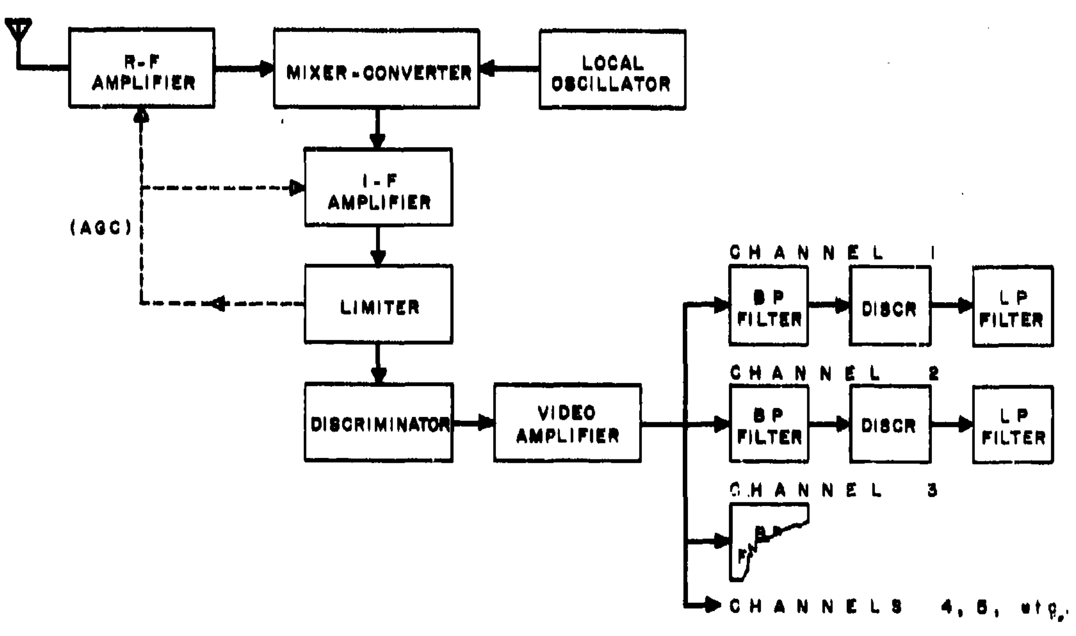


Figure 103. Typical FM/FM Receiver

The i-f amplifier and limiter are two additional stages of amplification. The limiter has the additional function of removing any amplitude-modulation components on the frequency-modulated signal, so that they will not be detected by the following discriminator circuit. The discriminator circuit then demodulates the i-f signal. The output (in the usual multichannel application) is the composite signal which was used to modulate the transmitter.

This output is amplified in a video amplifier and then applied through bandpass filters to individual subcarrier discriminators and filters for final extraction of the data signals that were being transmitted.

Chief considerations in r-f receiver design are stability and low-noise operation. Stability is implemented by the use of crystal-controlled oscillators, and by various schemes of automatic frequency control (afc compensates for drifting of the oscillator frequency after extended periods of operation). Low-noise operation is optimized, within the limits of a wide frequency response characteristic, by carefully tuning the r-f and i-f stages to reject all but the desired band of frequencies. Noise amplitude components also are minimized by the use of automatic gain control; this is obtained by feeding part of the negative bias from the limiter stage back to the r-f and i-f amplifiers, so that sharp amplitude rises at the limiter will reduce the output of the preceding stages.

3. Antennas

When a current (from a transmitter) alternates in a conductor at high radio frequencies, portions of the electric and magnetic fields that normally build up and collapse about the conductor (transmitting antenna) are radiated into space. The electromagnetic fields travel away from the antenna at the speed of light, and deliver a usable portion of power to a receiving antenna, which intercepts the energy at great distances from the transmitting antenna.

a. Basic Antenna Types

The basic antenna configuration is a half-wave dipole, consisting of two lengths of wire or metal tubing, each a quarter-wave long at the transmission frequency. A dipole set up in a vertical plane radiates equally in all directions. The electric wave from the antenna is in a vertical plane, as shown in figure 104, and the antenna is said to be vertically polarized. If the dipole is in a horizontal plane, it radiates a horizontally polarized electric wave, as shown in figure 104; its radiation is concentrated about a line perpendicular to the dipole, as shown in figure 105.

The operation of the half-wave dipole illustrates the basic principle of antennas. The wise variety of antenna configurations are designed for maximum effectiveness at various transmitting frequencies and ranges, and for particular patterns of radiation directivity.

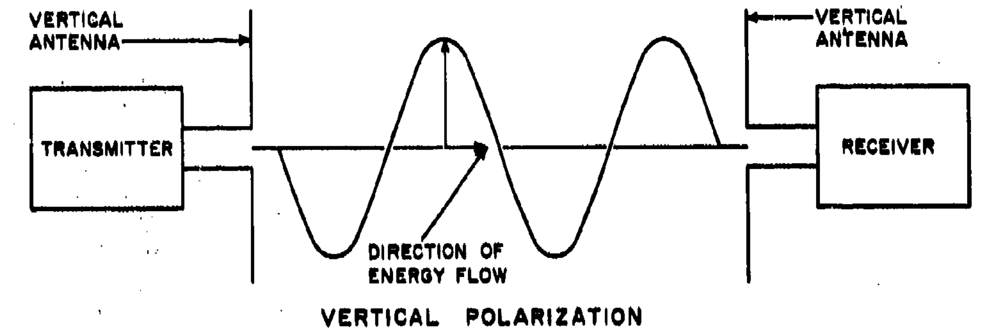


Figure 104. Polarization of Radiated Electric Waves

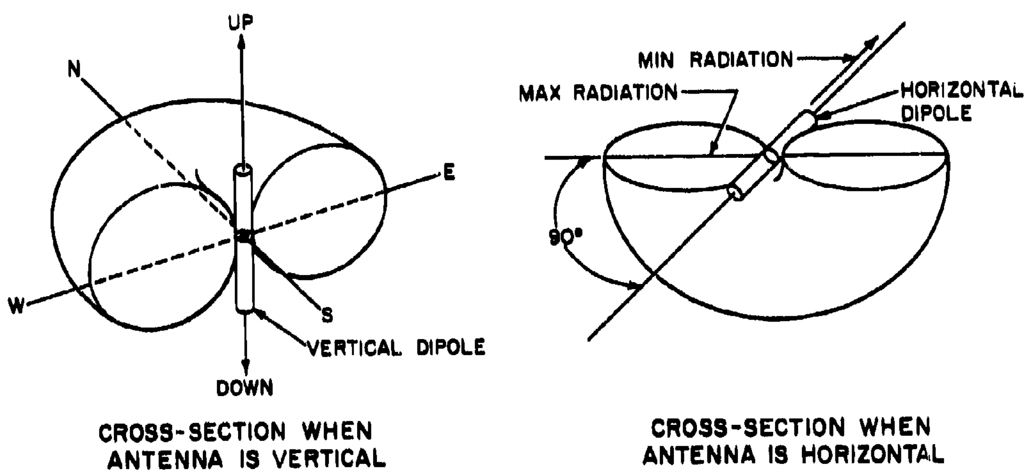
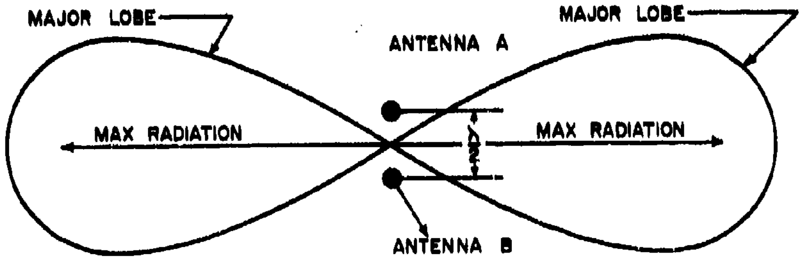


Figure 105. Radiation Pattern of a Dipole

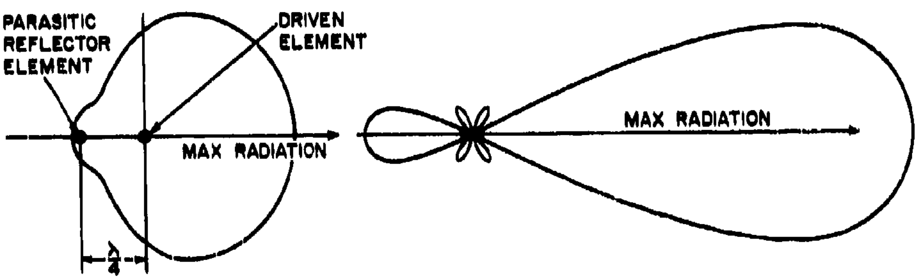
Multi-element vertical dipole-type antennas, for example, can exhibit horizontal radiation patterns quite different from the regular figure 8 shown in figure 105. A two-element antenna with two driven or excited in-phase elements radiates a similar figure-8 pattern, but with much narrower lobes (figure 106). A driven element with a parallel, unexcited element is unidirectional (figure 106). A Yagi antenna, which has one driven element and two or more parasitic elements called reflectors and directors, is essentially unidirectional, with a negligible minor lobe and a major lobe that is concentrated in a narrow beam (figure 106).

At microwave frequencies, the radio waves exhibit optical properties, and they can be focused, with reflecting elements of parabolic shape, into parallel lines forming a very narrow beam (figure 107). The excited element shown is again a half-wave dipole, with a parallel reflector to direct its radiation toward the parabolic reflector. The reflector may also be driven by a waveguide, which feeds the energy to the focal point of the parabola, from which it can be reflected against the parabolic surface.

Similar in reflective principle is the corner reflector, which consists of two conducting sheets that meet at an angle, with the exciting element in a plane which bisects the corner (figure 107). This type of antenna has somewhat higher gain than the parabolic type.



2 VERTICAL RADIATORS, IN PHASE



1 RADIATOR, 1 REFLECTOR

YAGI

Figure 106. Radiation Patterns of Multi-Element Antennas

Another highly directional antenna is the horn radiator. The horn is a travelling wave antenna, as opposed to the dipole types discussed above, which are normally standing-wave devices. In the dipole, the wavefront of the signal from the transmitter is reflected from the unterminated ends, setting up standing-wave patterns along the antenna, and making the fundamental radiation characteristic of the single-element antenna omnidirectional. The horn, however, is shaped so as to match the impedance of the waveguide or transmission line with the impedance of free space (377 ohms), so that a minimum of the energy is reflected after it leaves the transmitter. Thus, the radiation from the horn travels outward in the direction of the wave motion (figure 107).

b. Transmitting Antennas

In most telemetry applications, highly directive antennas are desirable. For transmission between aerospace vehicles and ground, however, aerodynamic considerations and the structural design of the vehicle may prevent the use of the most effective antenna. The relative position of the transmitter and the available antenna

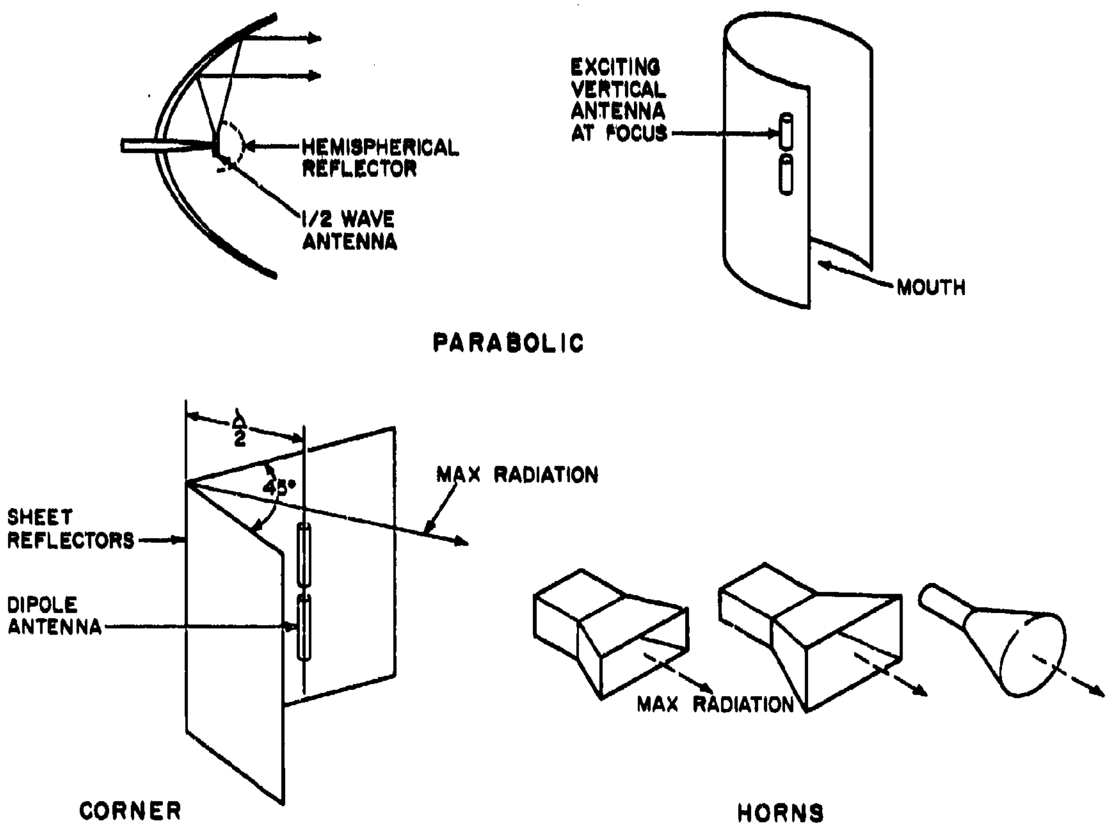


Figure 107. Reflector and Horn Antenna Configurations

locations may determine the antenna type. Wire radiators and reflector types such as that described above may be employed, or the antenna may be incorporated in some part of the vehicle structure.

Direct excitation of part of the vehicle body can sometimes be used to obtain an effective transmitting antenna, depending on the dimensions of and the ability to insulate the radiating part from the rest of the vehicle. An airplane's wings can be used as a dipole for long-wavelength, low-frequency transmission. Similarly, the nose cap or spike on high-velocity vehicles can sometimes be used as an antenna.

Under certain conditions, a notch or slot in a conducting surface can be used as dipole radiators. Such a notch can be cut in the trailing edge of a wing or tail fin, for example, and be fed by a coaxial line, or a slotted waveguide termination may be mounted flush with the surface of a vehicle.

c. Receiving Antennas

The principal receiving antennas used with missile and space telemetry systems are the parabolic dish, the helical beam array, and, occasionally, the dipole and Yagi types used with communications systems. The principal factor affecting the sensitivity and overall gain of receiving antenna arrays is size, and at ground receiving sites size is limited primarily by cost. If a stronger signal is needed in a given application, it usually is possible to make the antenna larger. As with transmitting antennas, receiving antennas should be located as close to receiving equipment as possible to minimize the length of transmission line used to link components.

With the directive transmission characteristics of most telemetry systems, antenna design factors such as polarization and orientation, which affect the sensitivity and gain of the antenna system, are extremely important. With parabolic dish antennas or helical beam arrays, some means of keeping the antenna pointed at the transmitting antenna, either manually or automatically (such as by slaving the receiving antenna to a tracking radar), must be provided.

To ensure a high signal-to-noise ratio, especially with long-range transmissions such as from space to ground, preamplifiers are sometimes used to boost the antenna output fed to the receivers. For this application, low-noise amplifiers of special design are employed; some new devices are parametric amplifiers, traveling wave tubes, and, most recently, ruby masers.

B. AM Systems

1. Principles and Techniques

Amplitude modulation is the variation of the amplitude of a carrier in accordance with the variations of an intelligence-carrying input signal. The input

signal normally is varying both in amplitude and in time (a frequency component), so the modulation is twofold: the carrier amplitude changes with the amplitude of the input signal, and it changes at a rate determined by the frequency of the input signal.

In practice, the carrier is a sinusoidal waveform generated at a radio-frequency rate (many thousand or million cycles per second); the input signal is a lower frequency waveform (less than 15,000 cycles per second), called an audio signal. The circuits that do the modulating normally are located in the transmitter of the carrier system. Figure 108 shows a block diagram of a typical amplitude-modulated radio transmitter.

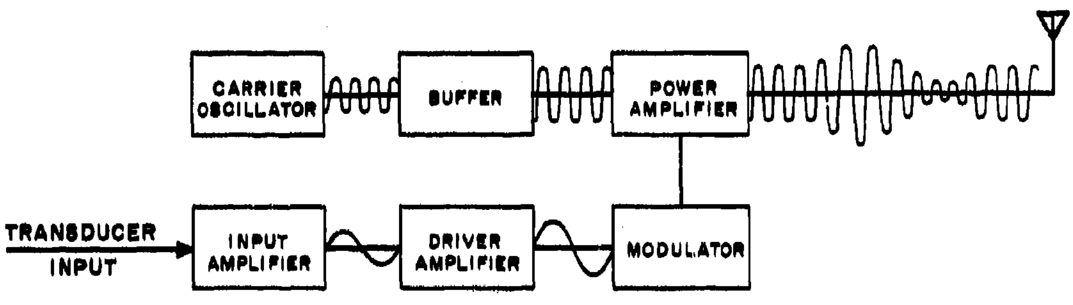


Figure 108. Typical Amplitude-Modulated Transmitter

The audio section contains two stages of input signal amplification and a modulator circuit. This modulator circuit normally is a coupling device, such as a transformer, which connects the amplified input to the power amplifier in such a way as to impress the audio variations upon the carrier. There are several techniques for accomplishing this, depending upon where in the power amplifier stage the audio is introduced. The following are the most common:

a. Plate Modulation. Here, the audio is added (or subtracted) from the output of the power amplifier.

b. Control Grid Modulation. The audio is applied to the control grid of the power amplifier, where it increases or decreases the bias on the amplifier, controlling its power output.

Plate modulation is known as a brute force technique; much more power is needed for modulation, since it is being added directly to the signal feeding the amplifier. Much less power is required for grid modulation, since it takes advantage of the amplification in the power output stage; however, the amount of modulation so achieved normally is less than half that possible with plate modulation.

2. Bandwidth Requirement

The wave produced by amplitude modulating a carrier with a low-frequency sine wave consists of three separate components: the carrier frequency and two sideband frequencies (see figure 109). The sidebands are the result of the sum and the difference of the carrier frequency and the modulating frequency. If a 100-kc carrier is amplitude modulated by a 100-cps signal, the upper sideband is 100 kc + 0.1 kc = 100.1 kc while the lower sideband is 100 kc - 0.1 kc = 99.9 kc.

Because an amplitude-modulated wave has sidebands on each side of the carrier frequency, this method of transmitting information requires that a band of frequencies be used rather than a single frequency. Consequently, if intelligence is transmitted using a maximum amplitude-modulating frequency of 5 kc, a bandwidth of 10 kc (5 kc on either side of the carrier frequency) is required. The sideband frequencies must be included in the transmission and reception of amplitude-modulated signals, since the intelligence is all contained in the sidebands.

3. Percentage of Modulation

The degree of modulation in an amplitude-modulated signal is expressed by the percentage of maximum departure from the normal value of the carrier. The effect of such a modulated wave, as measured by receiver response, is proportional to

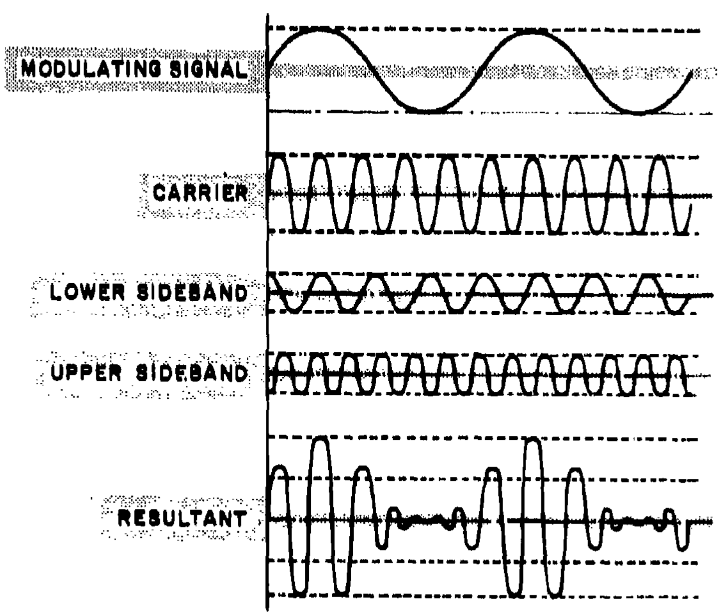


Figure 109. Waveforms Produced by Amplitude Modulation

the degree, or percentage, of modulation. The percentage of modulation may be determined by the equation:

$$\text{percentage of modulation} = \frac{E_{max} - E_{min}}{2E_{car}} \times 100$$

where E_{max} = the maximum instantaneous value of the modulated carrier
E_{min} = the minimum instantaneous value of the modulated carrier
E_{car} = the maximum instantaneous value of the unmodulated carrier.

Since the intelligence in an amplitude-modulated waveform is contained in the sidebands, as much power as possible should be placed into the sidebands. To accomplish this, a high percentage of modulation must be used; i.e., the amplitude of the modulated carrier should be varied as much as possible. When the amplitude of the carrier decreases completely to zero and increases to twice its normal unmodulated amplitude during the modulation cycle, the carrier is 100 percent modulated and the sidebands contain the maximum permissible amount of power (one-half the power contained in the carrier).

4. Demodulation

Demodulation is the process of recovering the intelligence from a modulated or an encoded wave. Demodulators (detectors) for an amplitude-modulated carrier produce voltages that vary in accordance with the amplitude of an amplitude-modulated wave. In other words, the action that takes place in a demodulator or a detector circuit is the reverse process of that which occurs in a modulator.

a. Frequency Conversion

In most receiving systems, the initial step towards converting the modulated carrier frequency to the desired information is to change the frequency of all received signals to a common intermediate frequency by mixing the received signal with a signal produced by a local oscillator within the receiver. The result of this mixing action (heterodyne action) is four different frequencies: the two originals, their sum, and their difference. The difference frequency is selected by appropriate band-pass filters and is used as the intermediate frequency. Regardless of the method of modulation used to impress intelligence upon the carrier, the information characteristics of the received signal still exist in the intermediate-frequency signal.

The principle of frequency conversion by heterodyne action can reduce any desired transmitted frequency within the receiver range to a common intermediate frequency. A received signal, therefore, is converted to the fixed intermediate frequency before the actual demodulation process takes place. The intermediate-frequency amplifiers in the receiver operate under uniformly optimum conditions throughout the receiver range; thus the intermediate frequency circuits can be made

uniformly selective, uniformly high in gain, and uniformly of satisfactory bandwidth to contain all of the desired sideband components associated with the modulated carrier signal.

b. Detection of Amplitude-Modulated Carrier

An amplitude-modulated carrier is detected by rectification and filtering. An amplitude-modulated wave is shown in A of figure 110. The modulating frequency varies the amplitude of both the positive and the negative half cycles of the modulated wave. The first function of a detector is to rectify the modulated wave in order to produce a series of radio-frequency pulses having amplitudes that vary in accordance with the modulating signal (B of figure 110). The radio-frequency pulses then are filtered out, leaving the desired information signal as shown in C of figure 110.

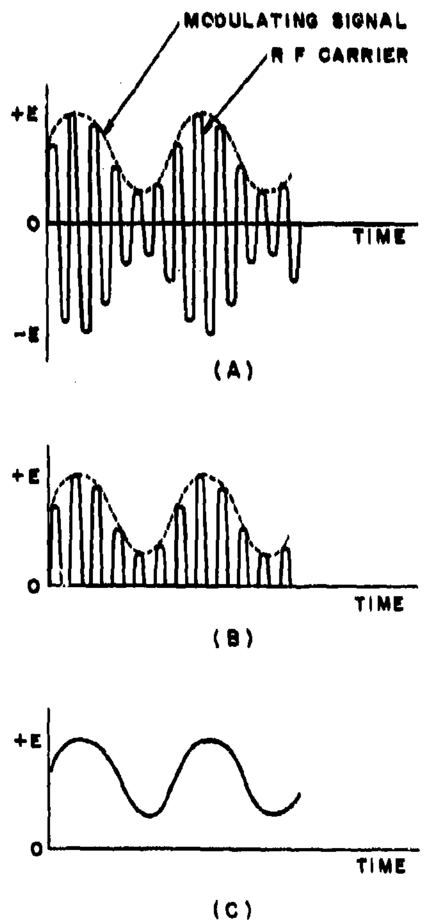


Figure 110. Detection of an Amplitude-Modulated Carrier

Several types of detectors are available for the demodulation of an amplitude-modulated carrier. One of the simplest and most widely used is the diode detector. It uses a simple diode for rectification and a low-pass resistance-capacitance filter to suppress the radio frequencies. Other types of detectors include the grid-leak detector and the plate detector. Although the grid-leak and plate detectors offer greater sensitivity than the diode detector, the diode detector is more linear and is capable of handling stronger signals without overloading.

C. FM Systems

1. Principles and Techniques

Frequency modulation is the variation of the frequency of a carrier in accordance with the variations of an intelligence-carrying input signal. Like amplitude modulation, it is a twofold variation: the amplitude of the input signal determines the extent of frequency change, and the frequency of the input signal determines the rate of frequency change. This preserves the amplitude-versus-time intelligence on the original input signal.

The circuits that do the modulating normally are associated with the transmitter of the carrier system. Frequency-modulated transmitters normally consist of a carrier oscillator, which generates a sinusoidal signal at a frequency well below that which is radiated by the antenna; one or more stages of frequency multiplication; the final power amplifier; and, of course, the modulation circuits. There are two methods of frequency modulating a carrier; direct-coupled modulation and variable-reactance modulation. Since phase modulation is related closely to frequency modulation, it is discussed as a third method. (Refer to figure 111.)

a. Direct-Coupled or Mechanical Modulation

The frequency of the carrier oscillator is determined by the inductive and capacitive elements in the tank circuit of the oscillator; changing these reactive elements changes (modulates) the frequency directly. In a direct-coupled modulator, a variable reactance device, such as a capacitor microphone, is placed directly in the tank circuit of the oscillator. In a monitoring system, this system has the advantage of connecting the mechanical-to-electrical transducer directly into the data-transmission link. Actually, direct microphone modulation seldom is used; but certain variable-capacitance transducers used as pressure or displacement sensors in physiological monitoring can be conveniently connected directly into the subcarrier oscillators of a multichannel system.

b. Variable-Reactance Modulation

In variable-reactance modulation, the modulating signal is applied to a tube or diode circuit whose reactance varies with changes in the modulating

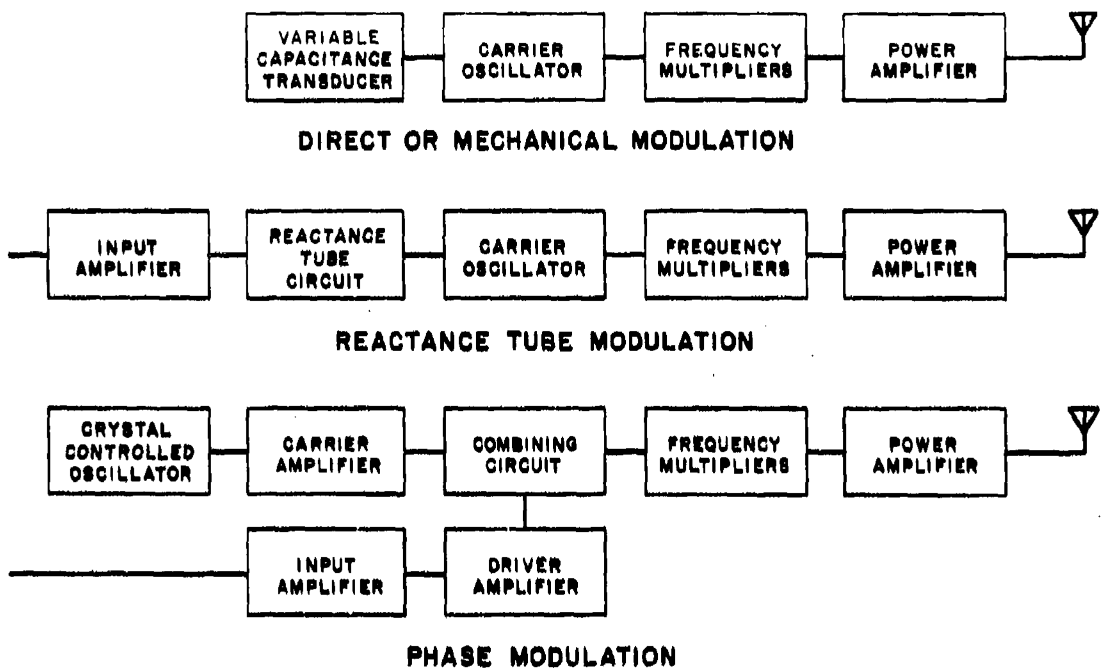


Figure 111. Principal Types of Frequency Modulation

signal. With the circuit connected across the tank circuit of the carrier oscillator, the frequency of the oscillator stage thereby varies with the input signal.

With no modulating signal in, the output of the tube (or diode) circuit, controlled by oscillation in the tank circuit, is a series of pulses at the basic carrier frequency. When a modulating input is applied across the grid of the tube, the plate current of the tube increases or decreases. The plate current drawn through the tank circuit acts like a change in capacitance, which varies the frequency of oscillation in the tank and the frequency of the pulses from the tube.

The basic frequency of the oscillator is a low radio frequency, and following stages of frequency multiplication (see figure 111) raise this to the high radio frequency that is fed into the power amplifier.

Other types of frequency modulation include inductance tube circuits for use with inductively tuned oscillators, and solid-state circuits using a varactor (a negatively biased diode). In a varactor circuit, the capacitive reactance of the stage is proportional to the signal voltage across the diode, and the modulating signal is applied to vary the bias.

c. Phase Modulation

When the instantaneous frequency of a carrier is varied by a modu-

lating signal, the result is frequency modulation. If the instantaneous phase of a carrier is varied by a modulating signal, phase modulation is obtained. However, varying a carrier's frequency also changes the instantaneous phase relationship of the modulated carrier frequency to the unmodulated carrier, and varying the phase of a carrier likewise changes the carrier's frequency. Thus, frequency modulation and phase modulation essentially are identical.

The prime distinction between frequency and phase modulation is the method by which the modulated wave is produced. Frequency modulation takes place at the source of the carrier signal; i.e., the operating frequency of the carrier oscillator is varied by the modulating signal. In phase modulation, the carrier is modulated in a combining stage (figure 111) following the carrier oscillator, i.e., the carrier is produced at a constant frequency and, after amplification and possible frequency multiplication, the modulating signal is applied. The end result of each of these methods is basically the same, and a phase-modulated wave can be received with equal facility by an FM receiver.

The combining stage is an amplifier whose output is controlled by the carrier and modulating signals both of which are applied to a voltage-divider network at the amplifier grid. The combined biasing causes the modulating signal to vary the phase of the amplifier output (rather than its amplitude which is the usual biasing effect).

2. Multichannel Considerations

Often when an FM system is employed, more than one channel of data is being transmitted, and these channels are multiplexed so that they share a single radio-frequency carrier. The multiplexing normally is accomplished by having each signal from each channel frequency modulate a subcarrier oscillator, and then combining their modulated outputs into a composite signal which is then used to frequency modulate the carrier oscillator in the radio-frequency section of the transmitter. This type of telemetry system is designated as FM/FM.

Figure 112 shows a typical modulation system for five channels of physiological data. The techniques of subcarrier modulation are the same as those described for carrier modulation, except that subcarrier oscillators normally operate at audio frequencies rather than at low radio frequencies.

a. Subcarrier Oscillators

Subcarrier oscillators are of two types: the reactance-controlled oscillator, which is linked directly to and modulated by a variable reactance transducer; and the voltage-controlled oscillator, which is modulated by the changing voltage from the input amplifier. There are many possible operating frequencies for these oscillators, but because of widespread FM/FM telemetry application, 18 basic

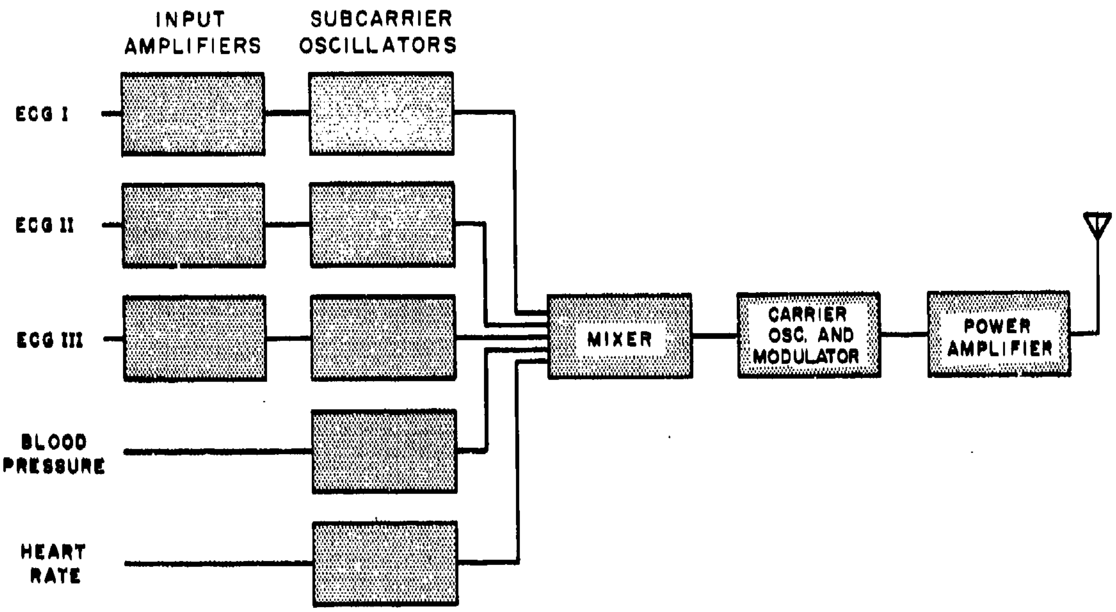


Figure 112. Five-Channel Telemetry Employing FM/FM

frequencies in the range between 400 and 70,000 cps have been standardized (ref. 54). Refer to table X for a listing of these subcarrier frequencies.

(1) Reactance-Controlled Oscillators

The reactance-controlled subcarrier oscillator operates in the same manner as the mechanical modulator. It essentially is an oscillator circuit with the transducer as one of its reactive elements. This type of oscillator must be calibrated with each transducer.

(2) Voltage-Controlled Oscillators

Most voltage-controlled oscillators are multivibrators whose square-wave output frequency is determined by the level of the input voltage. A stage of input amplification is normally required to raise the input signal to a level sufficient to deviate the multivibrator, and a low-pass filter is included at the output to reduce the sideband harmonics to prevent crosstalk with adjacent channels. (Older vacuum-tube voltage-controlled oscillators operated as an LC oscillator with a reactance modulator in parallel providing the frequency change with modulation.)

b. Mixing Circuits

The modulated outputs of the subcarriers in a multichannel system are combined in a resistor network to form a composite signal for modulation of the radio-

frequency carrier oscillator. If the level of the composite signal is inadequate to modulate the carrier oscillator, this stage also includes a mixer amplifier.

c. Carrier Modulation

Following subcarrier modulation and mixing, the composite signal then frequency modulates the radio-frequency carrier oscillator by one of the techniques described previously. It also is possible to take the FM composite signal from the mixing circuit and use it to amplitude-modulate a carrier, or to apply it as one input in pulse-type carrier modulation.

TABLE X. FM/FM SUBCARRIER FREQUENCIES

Channel	Subcarrier Center Frequency (Cps)	Deviation* Limits (Cps)	Useful Intelligence Frequency (Cps)
1	400	370 - 430	6
2	560	518 - 602	8
3	730	675 - 785	11
4	960	888 - 1032	14
5	1300	1202 - 1398	20
6	1700	1572 - 1828	25
7	2300	2127 - 2473	35
8	3000	2775 - 3225	45
9	3900	3607 - 4193	59
10	5400	4995 - 5805	81
11	7350	5799 - 7901	110
12	10,500	9712 - 11,288	160
13	14,500	13,412 - 15,588	220
14	22,000	20,350 - 23,650	330
15	30,000	27,750 - 32,250	450
16	40,000	37,000 - 43,000	600
17	52,500	48,560 - 56,440	790
18	70,000	64,750 - 75,250	1050

*Based on a maximum subcarrier deviation of 7.5 percent. Channels 14 thru 18 also may be operated with 15 percent deviation, provided certain adjacent channels are not in simultaneous use.

3. FM Sideband Structure

Whereas an amplitude-modulated wave contains one upper and one lower sideband for any modulating frequency, a frequency-modulated wave may contain many pairs of sideband frequencies. The number of significant sideband frequencies in a frequency-modulated signal is determined by the ratio of the carrier deviation frequency (the number of cps that the modulated carrier deviates from its normal unmodulated frequency as the result of a modulating signal being applied) to the modulating frequency. This ratio is referred to as the modulation index and may be expressed as follows:

$$M = \frac{F_d}{F_m}$$

where M = the modulation index
F_d = the frequency deviation of the FM carrier
F_m = the frequency of the modulating signal

The frequency and energy distribution for various values of modulation index are shown in figure 113. The center line represents the carrier, and the lines on each side of the carrier represent sideband frequencies. The lines to the right of the

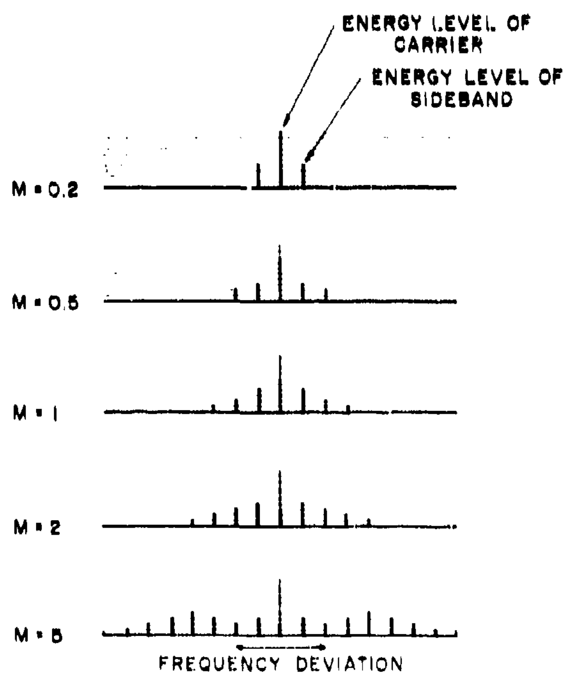


Figure 113. Frequency and Energy Distribution for Various Values of Modulation Index of an FM Wave

carrier represent the upper sideband frequencies, while those to the left of center represent the lower sideband frequencies. The lengths of the lines represent the energy levels of the various components.

No additional energy is supplied to an FM wave during modulation. Instead, the energy is redistributed. In figure 113 it is evident that the carrier actually is reduced when the modulation index exceeds 0.5. For values of $M = 2$ and $M = 5$, some of the sideband frequencies contain more energy than the carrier. When excessively large modulating signals are applied, the energy level of the carrier approaches zero.

For any given modulating frequency, the number of significant sidebands, and therefore the bandwidth requirements of the FM signal, depends entirely upon the frequency deviation of the FM carrier. The amount of carrier frequency deviation is determined by the amplitude of the modulating signal; therefore, it actually is the amplitude of the modulating signal, rather than its frequency, that determines the bandwidth requirements of an FM wave. The modulating frequency determines only the rate at which the frequency deviations take place. (In practice, of course, with multichannel telemetry standardization, the assigned space for each channel in the IRIG band is determined by the modulating frequency, and there is a direct relationship between frequency and bandwidth.)

4. Degree of Modulation

In an amplitude-modulation system, 100 percent modulation occurs when the amplitude of the modulated carrier varies between zero and twice its normal unmodulated amplitude. When this occurs, maximum power is being added to the sidebands and the system is operating at its maximum efficiency. However, the degree of modulation varies with the amplitude of the modulating signal, and, consequently, the system cannot be operated at maximum efficiency continuously.

In frequency modulation, 100 percent modulation has a different meaning. The modulating signal varies the frequency of the carrier only; therefore, the amplitude of the modulated carrier remains constant, and the oscillator stage operates at maximum efficiency continuously. A modulation of 100 percent means simply that the carrier is deviated in frequency by the full permissible amount in any given system or application.

5. Demodulation

To extract intelligence from an FM wave, the frequency variations must be converted into corresponding amplitude variations, which are detected in a manner similar to that used for AM detection. A simple method for doing this (known as slope detection) is to detune an ordinary filter circuit so that the carrier frequency of the FM signal is on the slope of the resonance curve (point B in figure 114). The amount

of frequency swing about the carrier varies with the amplitude of the modulating signal. When the carrier frequency of an FM signal falls on a sloping side of the response curve of a filter, the frequency variations of the FM signal are converted into equivalent amplitude variations because of the unequal responses above and below the carrier frequency (points A and C in figure 114). Thus, the output of the filter is made to vary in amplitude (in addition to its continuously changing frequency) at the information rate, and the resulting signal can be coupled to an AM detector to recover the intelligence.

Slope detection is seldom used because operation on one side of the resonance curve severely limits the deviation possible without distortion. A more common circuit is the frequency discriminator, which is tuned to the unmodulated carrier frequency; when the carrier swings (is modulated) to a higher frequency, the output amplitude increases in a positive direction, and when the carrier swings lower, the output amplitude increases in a negative direction, as can be seen on the curve of figure 114(B).

Figure 114(C) shows a simplified discriminator circuit for this function. FM to AM conversion is accomplished in the transformer circuit by combining the normal secondary voltages, on either side of the center-tap, with the primary voltage, applied through the capacitor. The relative phase of these voltage components is such

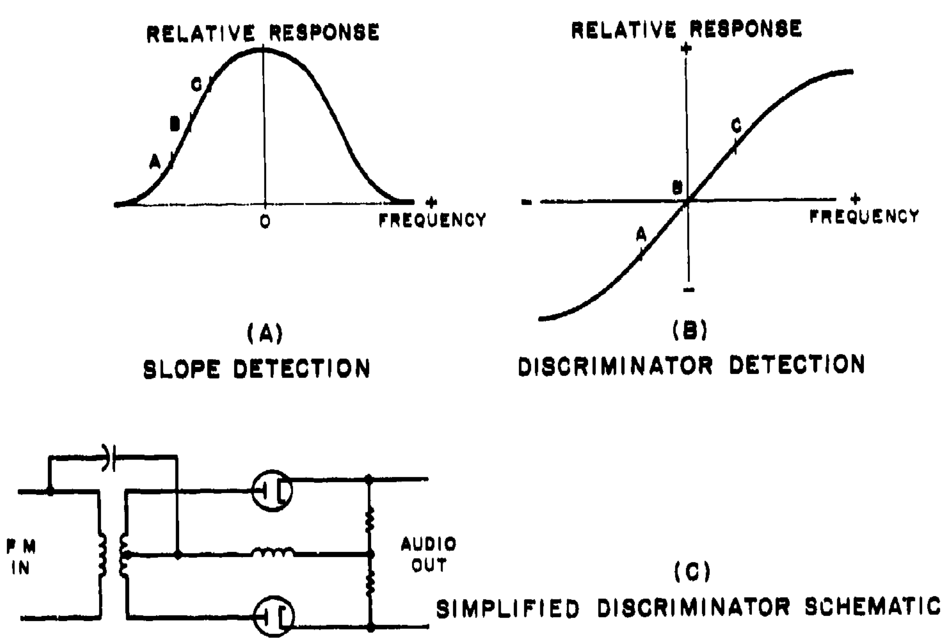


Figure 114. FM Demodulation, Including a Simplified Discriminator Circuit

that a change in the input (carrier frequency shifts) changes the voltages applied to the diodes in the secondary circuit, with the voltage through one diode increasing, and the voltage through the other diode decreasing. The difference between these two diode voltages, rectified, is the demodulated output of the discriminator.

As stated previously, one of the basic advantages of frequency modulation lies in its ability to minimize static and noise. This advantage is achieved by making the FM receiver insensitive to variations in signal amplitudes. To accomplish this, a limiter stage normally is used prior to the discriminator. The output of a limiter stage remains relatively constant in spite of varying signal amplitudes being applied to its input.

D. Pulse Systems

The various methods of pulse modulation are essentially sampling, multiplexing, and encoding techniques. The pulse circuits produce a composite signal that is used to modulate a subcarrier or carrier for r-f transmission.

To transmit analog data in sampled, pulse format, the sampling rate must be high relative to the rate of change of the sampled signal, if sufficient information is to be extracted for the analog signal to be restored at the receiving end of the link. The minimum sampling rate is between two and three samples for each cycle of a pure sinusoidal signal; for a nonsinusoidal waveform, the sampling rate is about two and a half times that of the highest frequency component in the waveform.

In a multiplexed system, the number of samples per second that can be obtained with a commutator must be divided among the channels being multiplexed, and the number of channels that can be multiplexed depends upon the frequency content of each channel.

1. Pulse-Amplitude Modulation (PAM)

Pulse amplitude modulation requires relatively simple equipment, since it requires only periodic switching to convert a continuous data signal into a series of narrow pulses with varying amplitudes. (See figure 92.) Data are seldom transmitted in PAM format because of magnitude domain limitations. But PAM is used for subcommutation in FM/FM telemetry, and it also can be used for the initial sampling of a signal prior to encoding the PPM, PDM, or PCM formats.

A common method for sampling continuous data for PAM transmission is to apply the input signals to the contacts of a rotary switch called a commutator. If the moving arm of the switch is rotated 30 times per second, it samples each cycle of a 10-cycle-per-second signal three times. If the inputs are applied to separate contacts of the commutator, they are sampled in sequence as the arm rotates, and multiplexing as well as sampling occurs automatically. The inputs normally are applied to every

other contact; the contacts between are connected together and a reference voltage applied to them to provide a base-line reference level between the sampling pulses. Figure 115 shows a typical PAM system configuration.

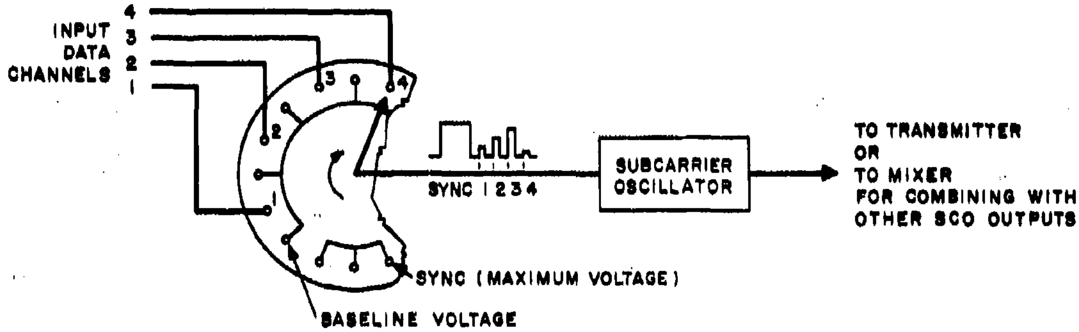


Figure 115. Typical Scheme for Commutated PAM

Sampling and multiplexing by this technique require that the transmitter and receiver be synchronized. This is done by including a sync signal at the beginning of each sequence of channel sampling (each frame). In a commutator, three adjacent contacts are tied together to obtain a sync pulse; this provides a pulse that is much wider than any of the data pulses. The sync signal normally is set at the level of maximum amplitude desired, and the various data channels then are adjusted to fall between that level and the interchannel baseline level.

The multiplexed output from the commutator then is applied in one of several ways to the radio-frequency carrier in the transmitter. It may be used to frequency modulate the carrier oscillator directly, it may be applied to a subcarrier oscillator and mixed with other channels in an FM/FM system, or it may be further encoded for pulse-duration modulation or pulse-code modulation.

Electromechanical commutators have their limitations: among them are a relatively slow sampling speed and irregularities caused by switching transients in the electromechanical contacts. Improved performance is possible with commutators using mercury jet switches. Electronic multiplexers are even better, though they increase overall circuit complexity. They are essentially logic circuits which sample data channels through electronic gates. Each channel is applied to a separate gating circuit, and the gating circuits (analogous to the commutator contacts) are closed momentarily, in sequence, by gating signals from a binary logic circuit driven by a master clock (oscillator). Electronic multiplexers are capable of extremely high-speed operation, so that more channels of higher frequency data can be multiplexed than are possible with an electromechanical commutator.

The signals are demodulated at the receiving end of a PAM data transmission system first by decommutation (or electronic demultiplexing), which separates

the multiplexed signals into individual pulse trains. The pulse trains then are processed through filter circuits, called interpolation filters, to restore the continuous information waveform.

2. Pulse-Duration Modulation (PDM)

Pulse-duration modulation was developed in the telemetry industry initially to overcome the disadvantages of PAM systems, particularly the noise transients produced at commutator switch contacts. What was done, essentially, was to take the PAM commutator output and apply it to a device called a keyer, which converted the pulses of equal duration and varying amplitude into pulses of equal amplitude but varying duration.

Figure 116 shows a block diagram of a typical PDM system. One set of contacts on a commutator samples the input data as in PAM, and a separate set of contacts provides a simultaneous set of trigger pulses. The trigger pulses activate a sawtooth generator, and the sawtooth output and the PAM signal are fed to a keyer circuit. There, the two signals are compared or combined to control a pulse generator, such as a Schmitt trigger, so that the pulses it generates vary in duration as a function of the amplitude of the PAM pulses.

The leading edge of each pulse is spaced equally, so that each data pulse is allotted the same amount of time in the sampling sequence. Therefore, the signal

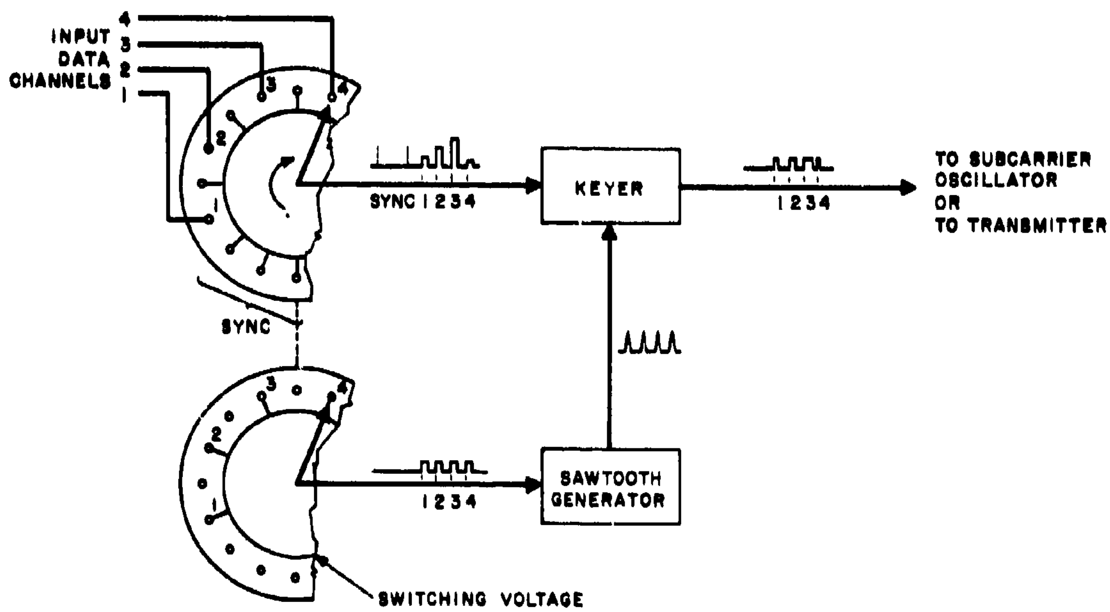


Figure 116. PDM System

levels into the commutator must be adjusted so that maximum amplitude deviation can be encompassed in the pulse time allotted, without overlapping the leading edge of the succeeding pulse and creating crosstalk between the channels.

As in PAM, sync is obtained by allotting two channels in each frame; however, instead of indicating sync with a long pulse, it is usually indicated by the absence of two consecutive pulses. One channel may also be allotted for zero reference, which is indicated by a narrow pulse of a predetermined width, which is shorter than the width of the narrowest (zero signal) data pulse.

3. Pulse-Position Modulation (PPM)

Pulse-position modulation is similar to PDM in that intelligence is encoded so that it may be extracted as a function of time or duration in a pulse train. In PDM, the duration of the pulse is the significant time interval; in PPM, the data pulses are reduced to a minimum, uniform width, and intelligence is contained in the time interval between the pulses and some time reference point in the sampling sequence.

In one type of PPM, the encoding process is essentially the same as PDM, except that after encoding, the pulse train is differentiated; the duration period of the pulse is eliminated and the leading and trailing (variable) edges are replaced by short pulses. In other words, a short pulse is produced at fixed intervals (this pulse represents the leading edge of the pulse in PDM), and a second short pulse (trailing pulse) follows at an interval determined by the instantaneous value of the sampled information signal. Thus, the relative position of the trailing pulses, in respect to the position of the fixed interval pulses, represents the instantaneous values of the information signal.

4. Pulse-Code Modulation (PCM)

In pulse-code modulation, the information is transmitted by means of a code of a finite number of symbols representing a finite number of possible values of the information at the time of sampling. A binary digital code normally is used; binary digits are represented in the code by the presence or absence of pulses in the assigned positions in a sequence of pulses. Figure 95 illustrates the relationship of a binary digital code to the decimal digital system. The presence of a pulse is indicated by the symbol 1 in the binary code, and the absence of a pulse is indicated by the symbol 0.

If each digit in the code has two possible values (0 and 1 as illustrated in figure 95), the number of different coded combinations or values is 2^M where M is the number of digits (pulses) in the code group. If a code is used in which each digit has three possible values, the number of different combinations or values is 3^M. In general, if there are x possible (and distinguishable) values of each pulse, the total number of values is x^M. The x^M points need not be spaced equally over the entire range of the information. It may be desirable, for example, to reduce the spacing in the neighborhood of zero in order to increase the accuracy of small signals.

A basic PCM decoder consists of a storage capacitor that is charged to a standard reference voltage at the instant preceding the reception of each code character. As the successive binary code elements that comprise a code character arrive, a subtraction circuit removes the capacitor's charge in proportion to the encoded value of the pulsed character. The total charge removed is delivered to the input of a low-pass filter. The result of a sequence of such operations is a reproduction of the original information signal.

5. Comparison of Pulse Techniques

Pulse-code modulation offers the best signal-to-noise ratio. As long as the pulses are above the threshold of noise level, fluctuation noise in a communication channel does not interfere with the recognition of the code groups. However, errors are introduced because only certain discrete values of the information can be transmitted by the code; i.e., in the process of coding, the signal must be quantized or digitalized to the nearest discrete value which can be transmitted by code. These errors are called quantization errors.

PCM methods are ideal for relaying purposes. As long as the carrier signals are above the noise threshold, neither noise nor errors are added by the individual links in the relay chain; errors are introduced only by the quantization before the initial transmission. For the same reason, PCM is ideal if the code groups are stored as such or fed into digital computing equipment. In this type of application, no additional quantization error is introduced at the receiving end of the system. PCM equipment, however, is larger and more complicated than the equipment for other forms of pulse modulation.

The limitations of PAM, a magnitude-domain technique, have been described previously; its principal use is as an adjunct to an FM system using other pulse-coding techniques.

PPM usually is preferred over PDM, especially when the time-multiplexed signal is used to amplitude modulate a transmitter carrier. Since only the variable edge of a PDM signal contains intelligence, transmitting the entire pulse results in power inefficiency. PPM improves this situation by eliminating the long duration period of the pulse and substituting a second short pulse in place of the trailing edge of the long pulse used in PDM.

E. Personal Telemetry Systems

The following paragraphs describe how the previously described fundamentals and techniques have been applied to the solution of a unique data transmission problem -- personal telemetry.

Personal telemetry is the monitoring of physiological or other data from a

freely moving, unencumbered human subject. It has many applications, including:

Basic research in physiology and psychophysiology, such as study of the reactions of athletes.

Medical technology, including diagnostic measurement and hospital patient monitoring.

Industrial safety, monitoring workers in dangerous or stressful environments.

Aerospace applications, such as the monitoring of an astronaut who must move and function in and around larger space vehicles of the future.

Personal telemetry is unique from other measurement situations in the strict requirement of subject freedom. All sensing, modifying, and transmitting devices must necessarily be attached to the subject. This places great restrictions on the size and weight of these devices, as well as on the power which they can consume (since the power supply too must be attached to the subject). With present techniques, this also compromises the information capacity, transmission range, and accuracy of the data transmission system.

1. Present Techniques Available

Data transmission necessarily must be accomplished with a wireless radio link.* Most commercially available systems employ miniaturized FM/FM components. The chief limitation of this approach has been accuracy and stability. The development of new components may result in much greater reliability.

Time division multiplexing methods, using pulse modulation, promise certain advantages for personal telemetry systems. One of the inherent advantages is that pulse circuitry can be designed around transistors operating nonlinearly, i.e., as two-state switching devices. In general, switching circuits permit less critical circuit tolerances, and more success has been achieved in developing microminiature integrated circuits with nonlinear operation, rather than linear.

Of the various pulse techniques, PPM and PDM are perhaps the most reliable for personal telemetry with the present state of the art of component design. PAM has certain advantages in simplicity, but with inherent magnitude domain limitations. PCM is probably the single most reliable coding technique, but it is presently

---

*Certain applications may warrant investigation of other basic transmission media, including other electromagnetic radiation such as light and infrared or sound (and ultrasound) waves.

impractical in this application because of component complexity. The following paragraphs describe a PDM system that was developed by the Aerospace Medical Division* to investigate the capability of a personal telemetry system built around relatively simple, reliable pulse circuits.

2. A Seven-Channel PDM System

Figure 117(A) shows a block diagram of a seven-channel PDM transmitter. Basically, it comprises seven signal modifying amplifiers, the r-f transmitter, and a pulse duration modulator, which includes the 1600-cps clock oscillator, a flip-flop binary counter, eight gating switches, and the Schmitt trigger, which produces the pulse signal shown in figure 117(B) to modulate the transmitter.

The 1600-cps output of the clock oscillator triggers the counting circuit, which consists of three flip-flop stages and a sawtooth generator. The flip-flop outputs are combined in a resistance matrix to produce sequential switching pulses for gating switches S1 through S8. Switches S1 through S7 gate the analog data signals with the sawtooth voltage. Switch S8 gates the sawtooth voltage plus a positive bias voltage to produce the synchronization gap between frames. The multiplexed outputs of the eight switches are delivered in sequence to the input of the Schmitt trigger. Since the Schmitt trigger turns "on" at a certain voltage level, the turn-on time will be a function of the channel information. The reset of the Schmitt trigger is derived from the return voltage of the sawtooth generator, and occurs therefore at constant time intervals. The leading edges of the pulses are produced from the reset of the Schmitt trigger, and the (modulated) trailing edges produced by the turn-on of the Schmitt trigger. The output of the Schmitt trigger (figure 117(B)) is used to frequency modulate the RF transmitter using a deviation of approximately 100 kc.

In the receiving unit, the signal from an FM receiver is fed to a PDM decoder. If only oscilloscope display of the analog information is required, the decoding circuitry is very simple and inexpensive. For recording on multi-stylus recorders, the decoder circuitry requires about the same amount of components (and expense) as standard FM discriminators.

In the system illustrated, the 1600-cps clock rate permits sampling each of seven channels 200 times per second. This limits the frequency response in each channel to 100 cycles per second, or an overall bandwidth of 0 to 700 cycles per second. The system permits paralleling of channels if higher frequency response is required, of course with a sacrifice in the number of individual channels. Several different combinations are possible in trading frequency response with number of information channels. Three examples other than the one described are:

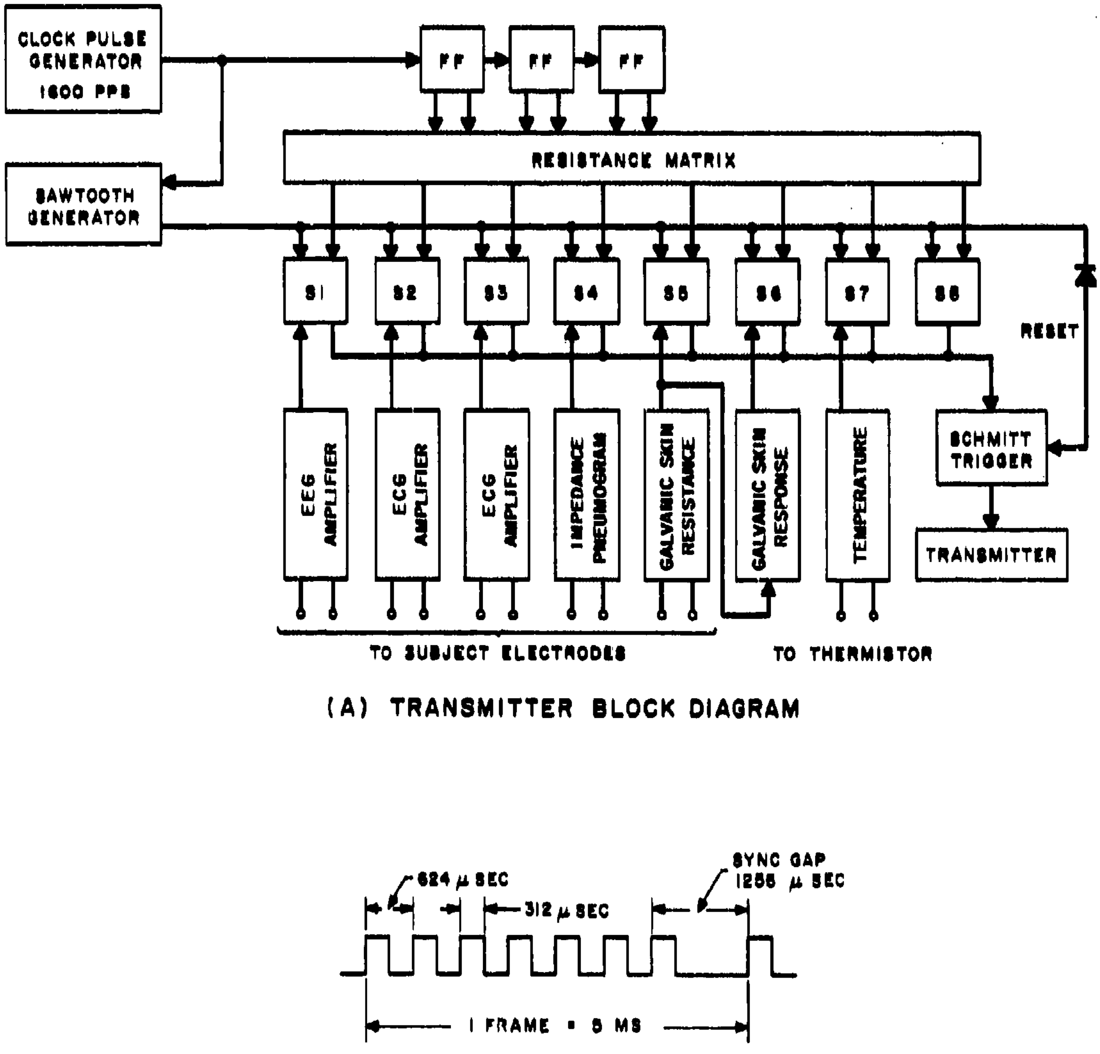
---

*6570th Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio.

Five channels with 100-cps response, and one channel with 200.
Two channels with 100-cps response, one with 200, and one with 300.
One channel with 100-cps response, and three with 200.

Since the frame repetition rate could be changed also, for different applications, the system is quite versatile.

Variation of the baseline in this system is less than 6 percent from full channel amplitude, over a temperature range of 32 to 82°F, with battery voltage



(A) TRANSMITTER BLOCK DIAGRAM

(B) PULSE FORMAT (UNMODULATED)

Figure 117. PDM Transmitter for Personal Telemetry

between 7.0 and 8.4 volts. Total power consumption is 68 milliwatts, including the seven signal conditioners. The power consumption of the multiplexing and transmitting unit alone is 35 milliwatts. With the use of a cross dipole receiving antenna and a receiver with 2 microvolt sensitivity, a useful transmitting range of 200 to 300 feet has been established.

F. Frequency Allocations

Communication links presently use frequencies ranging from 10 kc to 40,000 mc. By international agreement, the range has been subdivided into numerous bands, each identified with a particular service or group of services such as fixed, mobile, amateur, commercial broadcasting, etc. Each service has been assigned a plurality of nonadjacent bands throughout the frequency spectrum. Thus, the frequencies that can be used for physiological telemetering purposes are controlled by international regulations.

Frequency assignments in the radio spectrum are governed by factors such as the propagation characteristics at different parts of the frequency spectrum, transmitter power associated with a certain use (mobile stations almost always have lower power available than fixed stations), the limiting size of an antenna array (only small antennas are compatible with aircraft and space vehicles), the range desired, an estimate of interference to be encountered (both man-made and natural), the erratic qualities of the upper atmosphere as a factor in propagation, and the feasibility of channel sharing. The ultimate goal in a frequency assignment is a maximum ratio of signal strength to background noise, which is the best guarantee of successful operation.

Several factors affect frequency assignments for telemetering physiological data. The transmission of data from airborne, or even spaceborne, transmitters to remote receivers on Earth makes it necessary to consider the propagation characteristics between space and Earth. Only certain portions of the radio spectrum can even be considered for space-to-Earth radio links, because the lower layers of troposphere and upper layers of ionosphere are frequency selective, allowing some signals to pass through a "spectrum window" and reflecting, refracting, or absorbing other signals. Since this "spectrum window" covers the frequency range of approximately 10 mc to 10,000 mc, telemetering assignments are limited primarily to frequencies within this range.

The three frequency bands allocated primarily for telemetering purposes are 216 to 260 mc, 1435 to 1535 mc, and 2200 to 2300 mc. Special telemetering frequency assignments also are made by the Federal Communications Commission. An example is the satellite telemetering transmitters operating in a narrow band between 108 and 109 mc.

1. 216- to 260-Mc Band

Channel spacing in the 216- to 225-mc portion of this band is based on 0.5-mc separation on the integral and one-half megacycle channels (216.0 mc, 216.5 mc, 217 mc, etc). Assignments are made so that other established services are not interfered with. In the 225- to 260-mc portion of the band, a total of forty-four 500-kc channels are allocated on a protected basis until 1 January 1970. The maximum permissible r-f deviation of the FM transmitters is plus or minus 125 kc, and the transmitted r-f carrier must at all times be within 0.01 percent of the assigned carrier frequency. Although a maximum of 100 watts is allowed, the transmitted power should be held to a minimum consistent with satisfactory service. All harmonics and other radiated spurious signals must be at a level at least 60 db below the fundamental radiated power level.

2. 1435- to 1535-Mc Band

The 1435- to 1485-mc band is reserved primarily for telemetering applications in connection with aeronautical flight testing of manned aircraft. The range from 1486 to 1535 mc is reserved primarily for aeronautical flight testing of missiles and space vehicles.

Channels are spaced in increments of 1 mc, and the transmitter r-f carrier, including drift and all other variables, must be within 0.005 percent of the assigned carrier frequency. The transmitter output power is dictated by each particular application; however, it should never be more than is absolutely necessary.

3. 2200- to 2300-Mc Band

Channels in the 2200- to 2300-mc band are spaced in increments of 1 mc. The maximum allowable transmitter frequency tolerance is 0.005 percent of the assigned carrier frequency, and the amount of transmitted power is determined by each individual application.

Section VI

DATA PROCESSING EQUIPMENT

INTRODUCTION

1. Classification of Data Processing

The last steps in the chain of physiological monitoring are the analysis of the data and the drawing of conclusions. In many instances, direct inspection of the signal is all that is necessary to accomplish these last steps. In other cases, simple manual operations such as the plotting of body temperature over a long time span, are needed to reveal significant relationships more clearly. But as the data become more complex, automatic data processing may be the only practicable method.

Data processing involves some change in the form of input data to extract more meaning. The change may involve the simple regrouping or rearrangement of the data, or the employment of sophisticated mathematical techniques. The meaning may be hidden in some form not readily interpreted, or it may have meaning only when it is related to other information inputs. Data processing includes such processes as quantification, codification, transformation, and comparison. Each of these processes may be performed manually, such as using a slide rule, hand collating and graph plotting, or may be performed automatically, using card sorters, analog computers, or digital computers. If automatic techniques are used, care must be taken that the experimenter understands both the capabilities and the limitations of the equipment used.

Data processing is a general term applied to special operations performed on certain signals. It is difficult to draw a definite line between operations classed as signal modification and those classed as data processing, especially where the processor is an analog computer. In many cases the decision is arbitrary. However, the term data processing generally applies to the more complex operations, particularly those involving the accurate and rapid interpretation of several channels of data. Automatic data processing equipment, of which a computer is the central unit, not only records and stores data, but performs arithmetic operations as well as "logic" and "decision" functions.

Undoubtedly, computer techniques are not being exploited fully at present because of lack of training on the part of users, a condition that can be corrected with education. A more serious problem is the tendency to overestimate the capabilities of computers, fostered by the popular press and publicity releases. There is more to the problem of analyzing large quantities of experimental data than just feeding the data to a computer and expecting the computer to provide all mental processes and supply

all answers. This misleading idea has in the past often resulted in the accumulation of boxcars full of machine produced computations with no real answers to problems. It is most important to realize that the computer used for data processing can do nothing that it has not explicitly and in detail been told to do. It cannot perform thinking processes beyond the conceptual and programming ability of some human. Certainly, after once being instructed or programmed, the machine can perform a gamut of operations in an incredibly short time.

Large computers possess certain capabilities far beyond those of the human individuals. The following characteristics illustrate some computer advantages:

1. Infallible memory. The computer possesses the ability, when properly instructed, to retain and relate data for indefinite periods. It does not forget or overlook.

2. High computational speed. The computer can perform complex mathematical and logical operations at rates far faster than the human.

3. Indefinite operating capability. The computer does not get fatigued and can operate indefinitely without rest or error.

4. Not single minded. Parallel operations can be performed simultaneously, unlike the human mind that cannot place attention upon more than one factor at a time.

II. Classes of Data Processing Functions

A. Manually Operated Devices

Manually operated devices rely on both the skill and accuracy of the human operator to transcribe experimental data into a form that has some significance when interpreted in the framework of the experiment from which it derived. The data may be presented to the operator in many ways: visual displays by means of meters, recorders, oscilloscopes, or photographs; visual printout displays including typewritten, handwritten, or teletyped data; or aural indications such as warning tones. In each case it is the responsibility of the operator to interpret the indication correctly, quantify it by some comparative technique, either formal or informal, and transcribe the data. The operator may then perform various mathematical operations on the data to derive meaning from the raw data.

The simplest mathematical operations involve basic arithmetical functions, such as addition, subtraction, multiplication, division, squaring, and square root extraction. These processes may be performed with the aid of slide rules, desk top calculators, or tabulated tables of functions of the data (such as square root values)

available from published sources. If trends in the monitored data are of interest, then graphing techniques may supply the desired form of display.

1. The Slide Rule

The slide rule may be considered as an analog device; the positions of numbers on scales represent the logarithm of the number, so that multiplication is performed by the addition of scale distances, which in turn might represent the values of a parameter. Its computational speed is determined both by how conveniently the scales are arranged on the slide rule and by the mental and manual dexterity of the operator. The accuracy of the slide rule is limited by the resolution provided between adjacent markings on the scales and by the operator's judgment in interpolating values occurring between marked numbers. Extra long slide rules and circular slide rules, two variations of the standard 10-inch rule, are used to increase the length of the scales and thus increase the resolution and accuracy of the device.

The most common computations made on the slide rule are multiplication and division. In addition, other scales arranged in various manners allow the following functions of a number to be determined: the inverse, sine, tangent, logarithm, square, cube, square root, cube root, hyperbolic sine, hyperbolic tangent, and exponential.

2. The Desk Calculator

The desk calculator performs addition and subtraction by counting mechanically. The counters are cogs on gears which are arranged so that decimal arithmetic is performed. The numbers to be added or subtracted are entered into the machine by a keyboard, a motor drives the gearing, and the arithmetic answer is either displayed on a set of digits inscribed on one of the gear sets or is printed out on a paper tape. By extra electromechanical features included in the desk calculator, multiplication and division are also performed, which are accomplished by successive additions and subtractions. Other functions including square root derivation may also be used in data analysis if special tables are available.

The desk calculator can be accurate to as many decimal places as desired, if size is no factor. Since the device is digital in nature, and has only discrete values with no interpolative operation necessary, its correct operation is insured, if the correct information is entered by the human operator. The speed of the machine allows addition in less than a second and multiplication and division in several seconds.

3. Manual Graphing Techniques

By entering calculated or observed data on a graph paper ruled in some convenient pattern, trends in data may be observed and algebraic and transcendental functions of the data may be computed. There are many varieties of graph paper, the most common being the linear-rectilinear pattern. On this type of paper the plotting

of data versus another parameter (often time) shows the deviation from linearity quite readily. Another ruled pattern, the polar coordinate paper, is better suited for the graphing of data like electrical potential distributions whose relationships are more easily interpreted in this display. Semilog and logarithmic-logarithmic (log-log) paper contain scales where one or both dimensions are ruled in terms of the logarithm of the data being charted. One-coordinate log paper (semilog) permits the compression of one parameter, such as time or frequency, while still allowing a full expansion of a dependent variable, such as amplitude. Log-log paper is extremely useful for determining whether the function being analyzed follows an exponential dependence; if it does, a straight line is obtained for the data plotted. Other special graph patterns available are probability function-vs-linear scales, probability function-vs-logarithmic scales, reciprocal absolute temperature-vs-linear scales, hyperbolic scales, and trigonometric scales.*

The accuracy of graphing techniques depends on the choice of paper and the care taken in plotting the values. In general, graph plotting is a time-consuming, tedious process. However, the plotted experimental data indicate the variation of a physiological function better than any tabular or numerical representation.

B. Automatic Devices

Automatic tools of data analysis include computers, both digital and analog, and card sorters. The digital computer is a relatively complex, expensive, but extremely fast device that operates with quantities that are digitized (not of a continuous nature but composed of discrete values). All operations, even the most complex, are based on simple arithmetic processes - addition and subtraction. The precision of digital computers is limited only by the number of digital places available. They must have number storage (memory) capability for all but the simplest operations.

Analog computers may be very simple devices, consisting of a few electronic amplifiers and associated components, or they may be rather complex, involving an entire roomful of equipment. This type of computer generally needs no discrete memory capability; various complex operations are possible, including adding, subtracting, multiplying, squaring, integrating, and differentiating. The accuracy and precision of the analog computer are limited by the components available in its construction and the noise factors associated with electronic equipment. A fixed percent of accuracy describes the typical device. The analog computer is limited in its operational speed and, therefore, the frequency range of data input it can handle. Ten cycles per second is representative of analog computers using electromechanical devices in its operations, and 200 cycles per second is typical of all-electronic types.

---

*Keuffel & Esser Co., New York.

Card sorting equipment also manipulates information reduced to a digital form. While no computational processes are associated with this type of automatic equipment, the fast speed and ease in the handling of data reduced to card form make this form of data processing very useful.

1. Digital Computers

Digital computers, used increasingly in the medical sciences, are extremely important to physiological monitoring. Combined with appropriate peripheral equipment, the digital computer system enables the experimenter to collect, sort, correlate, compare, and analyze volumes of data available from the monitoring chain and to draw conclusions from the processed data, all done in minutes which otherwise would have taken weeks or years without automatic equipment.

The extreme speed of the digital computer results from its use of electronic circuits whose operations are capable of being performed millions of times a second. The flexibility of the computer and its power to make simple, logical decisions stem from its ability to compare data and follow a course of action dependent on the results of the comparison.

The digital computer may be operated either on line or off line; i.e., the computer may be connected so that the input devices are receiving data as the experiment occurs and the computer computes in real time (during the experiment), or the input to the computer is data transcribed from a storage medium, such as magnetic tape or punched cards, and the computation is carried out at some convenient time after the experimental data are gathered. In either case, the data presented to the computer must be in a digital form requiring an analog-to-digital conversion process prior to the input to the computer. The output of the computer also may be stored for future use or analysis, or it may be connected on line to control the progress of the experiment.

The digital computer can provide as much accuracy as the user or designer desires. Since it is a digital configuration, the more places (powers of two) available, the more accurate is the represented number. (See page 204 for a discussion on digital notation.) To maintain computer accuracy and to detect any errors arising because of equipment malfunction, error-detection circuitry is incorporated within the logical design of the digital computer.

The speed at which computations are performed in a large scale digital computer depends on the design of the computer and the type of operation performed. Since most computer operations are performed by successive additions and subtractions, the time necessary for one addition becomes the basic timing element in the computer. In some high speed computers, nearly one million additions may be performed per second.

2. Analog Computers

An analog computer is a configuration of components used to manipulate input data so that a desired mathematical function of the input is obtained at the output of the computer. Usually the analog computer deals with electrical inputs and outputs, but mechanical analog devices have been built and used. The voltages generated and manipulated may be actual voltages generated and present in an electrical system, or they may be analogs of mechanical, hydraulic, or other physical systems, with voltages representing forces or pressures. An analog computer has two main functions in an experimental situation. First, if the physics of an experimental situation can be reduced to a stated equation, the analog computer usually can solve the equation; and, second, it is used for processing of experimental data to derive additional information. (A third use, as a simulator of the system under study, is not of much interest in the field of physiological monitoring.)

The principle of operation of the electrical analog computer is based upon the ability of an operational amplifier and associated components (discussed in section II) to add, to multiply and divide by a constant, to integrate, and to differentiate. A mechanical analog computer depends on gearing, differential and regular, and mechanical disc integrators for its operation. By properly combining its functions, and introducing additional specialized operations, the analog computer becomes a very versatile instrument.

3. Card Sorters

Card sorters are one of the oldest forms of automatic data-processing equipment. The card sorter provides a well integrated and automatic mehtod of categorizing and filing information for fast and convenient sorting and retrieval. One standard Hollerith card (IBM card type) has 80 columns, and each column has 12 rows (see figure 118). The first 10 rows have decimal digit significance (0-9), while the remaining two rows are used for alphabetical and special symbol representations. Each of the 80 columns can be punched to represent one number letter, or special symbol. Actually, the punched card is a form of memory device with a special access means.

Once information is transcribed onto the punch card (with special punching typewriters), the cards are stacked and filed for future reference. To extract information from the cards, the cards are put into a sorting machine where each card is examined for a special punched sequence corresponding to the information being searched. By an electromechanical arrangement, all of the cards containing the information of interest are extracted from the card stack. Punched cards also are used as input and output media for digital computers.

Most punch typewriters also imprint the symbol punched in the column above the punched symbol, permitting simple proofreading of the card to ascertain the accuracy of transcription. Another method of verifying the punching accuracy involves

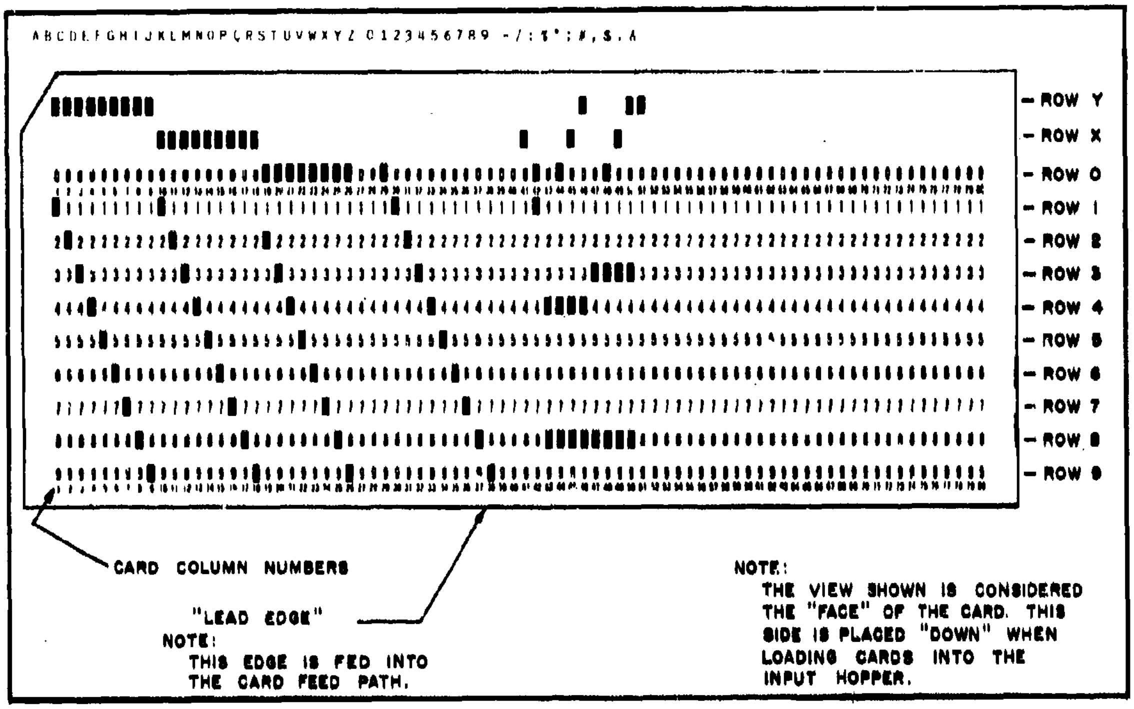


Figure 118. 80-Column Punched Card

punching the card a second time and noting any additional punched rows. If the accuracy of the punch mechanism itself is doubted, some card punching machines read the punch card immediately after punching and compare the required signal with that received from the readback. Any discrepancy activates an error signal.

III. Determinants for Data Processing Method

A. Quantity of Data

The factor that probably most influences the choice of data processing technique is the quantity of data or data analysis involved. If relatively few discrete values of data are obtained and a simple indication of variable change is all that is of interest, then manual graphing is indicated. If the quantity of data is so extensive that it will be impractical for a human operator to process the data and interpret results, then automatic data processing is used.

B. Speed of Processing

How quickly the processed data are needed is a determinant as to whether or not automatic data processing is used. If the data to be analyzed consist of involved, lengthly, sequential steps, automatic techniques may be necessary because the computer has the advantage of having all data practically instantly accessible, while a

human operator or group of operators is relatively inefficient in coordinating data segments.

C. Accuracy

The digital computer, when operating properly, may be considered error free, which unfortunately, is an attribute the normal human lacks. Not only can the computer perform arithmetical and logical functions with an extremely small probability of making an error, but it can be programmed to recognize certain types of errors in data introduction by various external sources. Such systems may be instructed to indicate not only the presence of the error, but at what point the error is introduced and which particular equipment malfunctioned.

D. Versatility

The results obtained from a single set of input data may be quite varied, depending on the mathematical or logical formats used with the data. Because of the speed and versatility of the computer, multiple analytic techniques can be attempted with automatic data processing that otherwise would involve prohibitive time and cost requirements. Such mathematical techniques as cross-correlation and autocorrelation are now being used to analyze experimental results that, previous to the advent of automatic techniques, were not even attempted because of the considerable expense and time required.

DIGITAL COMPUTERS

The digital computer is composed of four major elements: the input/output equipment, the arithmetic unit, the memory, and the control unit. The interaction of these elements allows the computer to perform various arithmetic operations, and, by proper programming, to perform various logical and control functions. The input/output devices are used for entering the desired program and data into the computer and extracting the results of the various computations from the computer. The arithmetic unit performs the operations. The memory retains stored information necessary for the performance of the various computer operations, and the control unit is the part of the computer where the operations are organized.

1. Digital Operation

The digital computer operates on numbers; numbers are discrete quantifications of physical phenomena. The most familiar number system is based on a decimal system, with the quantities 0 through 9 representative of increasingly ascending numerical values. To represent 10 distinct numbers electrically, the computer would have to be able to recognize 10 discrete values of voltage, which would present a rather difficult task. Instead, a number system is picked which is more suitable for electrical representation. By using a "0" or "1" state for a digit (binary), the computer has but

to recognize the presence or absence of a voltage. Thus counting is done by on-off sequences of voltages or pulses.

A character is composed of a sequence of the binary digits (bits). The binary code used is a positional code; i.e., a bit has a value according to where the bit is positioned in the word. This is the same convention used in the familiar decimal system. In the decimal system the digit to the extreme right in a number has the value between 0 and 9. A digit appearing in the next column to the left has a value between 10 and 90, and so on. This worth of the digit is based on powers of the system base or radix (10 in this case). Thus the four-digit number 3659, in this positional representation with base 10 may be expressed as

$$(3 \times 10^3) + (6 \times 10^2) + (5 \times 10^1) + (9 \times 10^0).$$

Each place has 10 possible values, with the digits 0-9 being the multipliers of the powers of 10. In the binary system the same configuration holds true, except that instead of a column signifying a power of 10, it is a power of 2. Thus the value 45 in binary notation is expressed as

$$(1 \times 2^5) + (0 \times 2^4) + (1 \times 2^3) + (1 \times 2^2) + (0 \times 2^1) + (1 \times 2^0).$$

In normal binary notation this would be written as 101101 (32 + 0 + 8 + 4 + 0 + 1). Obviously, the binary notation requires more digits to express a value than the decimal system, but the important feature of the binary system is that it can be represented electrically by having a "1" represented by a circuit that is on, and a "0" by a circuit that is off (or vice versa).

A convenient way of using binary notation is the 8-4-2-1 or binary-coded-decimal notation. Table XI lists the different notations for numbers, including decimal, straight binary, and binary-coded-decimal. It can be seen that in binary-coded decimal (BCD) each group of four binary digits represents one decimal digit, with both binary and decimal positional worth retained. In each group of four bits there are sixteen possible combinations of bit arrangement, with five unused combinations (decimal 10-15). These spare notations may be used to represent letters of the alphabet.

The on-off representation of numbers extends to all phases of the computer. A memory device must either be in one state or the other, also representing a "0" or "1" digit. Full numbers (or characters) are composed of a sequence of individual digits. A sequence of characters is called a word. The word is the unit transported at one time from one section of the computer to another. Each computer manufacturer places different requirements or specifications on the make-up of a word. The usual word length is about 24 or more characters.

A word may be transferred from one part of the computer to another in either a serial or parallel manner. In a serial transfer, one binary digit (bit) at a time is

transferred, so that it takes as many time sequences to transfer a word as there are bits in the word. A parallel transfer; however, transfers all the bits in a word at the same time over as many paths as there are bits. Therefore, a serial transfer requires more time but only one transfer path, while a parallel transfer requires less time but multiple transfer paths. Transfers of information go from memory locations to registers, or from registers to memory locations. There are additional paths between input/output devices and registers.

TABLE XI. DIGITAL REPRESENTATION OF NUMBERS

Decimal	Straight Binary	Binary-Coded Decimal
00	000000	0000 0000 0000
01	000001	0000 0000 0001
02	000010	0000 0000 0010
03	000011	0000 0000 0011
04	000100	0000 0000 0100
05	000101	0000 0000 0101
06	000110	0000 0000 0110
07	000111	0000 0000 0111
08	001000	0000 0000 1000
09	001001	0000 0000 1001
10	001010	0000 0001 0000
11	001011	0000 0001 0001
12	001100	0000 0001 0010
13	001101	0000 0001 0011
20	010100	0000 0010 0000
21	010101	0000 0010 0001
60	111100	0000 0110 0000
61	111101	0000 0110 0001
62	111110	0000 0110 0010
63	111111	0000 0110 0011

Figure 119 illustrates the various information and control paths in a large scale computer. As can be seen, the control unit is in charge of the computer operation once the computer starts. The control unit in turn is controlled by the computer instructions or program presented by the computer programmer in the form of detailed step-by-step instructions.

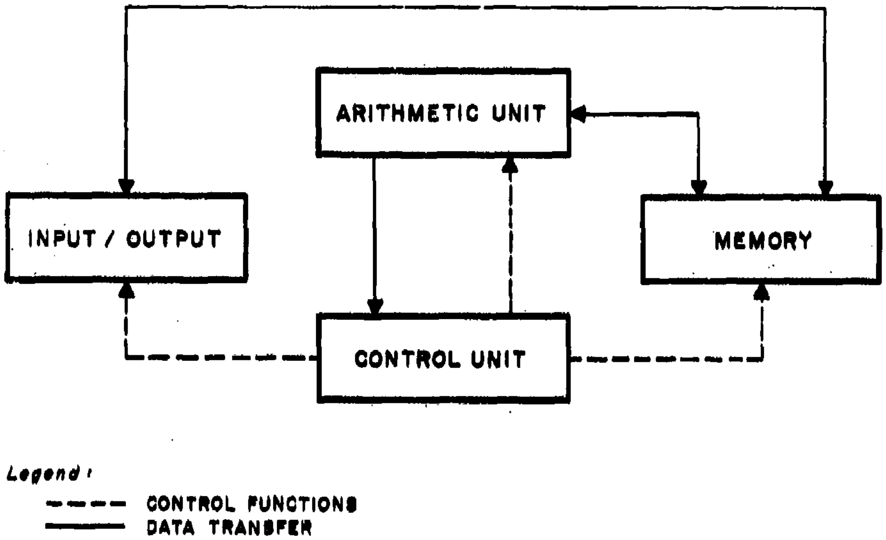


Figure 119. Block Diagram of Digital Computer Organization

II. Components

A. Input Devices

The digital computer requires input data and instructions written in a digital form, composed of on-off, "1" "0" characters. Once the data and information are entered into the computer, the machine operates automatically. One problem in the design of input devices is matching the extremely high speed of the electronic computing circuitry to the relatively slow electromechanical action of the input devices. Of necessity, a buffer (or storage stage) is connected between the input stage and the rest of the computer. The buffer receives and stores the relatively slow input data until it is economical to feed the data into the computer.

1. Magnetic Tape

Magnetic tape is the usual method for feeding information into a computer. The speed with which it can feed data approaches but not quite reaches the speed of computer operation. An average transfer rate from a magnetic tape station is about 50, 000 characters per second.

The operation of a magnetic tape station is similar to that discussed in Section IV. D-c pulses of information are recorded in parallel, with each track representing one bit of information. The number of tracks on the tape is determined by the number of bits per character and the type of computing system used. One system in general use, the RCA 501, has seven bits per character (six bits actually comprise the character and the seventh bit is for error checking), and an eighth bit is recorded for

timing purposes. Each of these eight bits is recorded on two separate tracks to avoid the loss of information because of malfunction in any one head element or an irregularity in the coating of the magnetic tape. Thus 16 tracks, with each one responsible for one of the eight bits or pulses, comprise the character. The character from the tape is retrieved in parallel, requiring eight amplifier channels.*

The tape mechanism can run the tape forward or reverse its direction past the heads used for recording and sensing. The computer controls the motion of the tape. Figure 120 illustrates a tape station used for a computer system.

The magnetic tape station is used for entering input data and instructions, and as a memory device for the computer, handling the overflow from the high speed memory in the computer. A third use is as a "scratch pad" for the computer. When calculations are to be performed and the results are being used for additional calculations, the magnetic tape medium is used for this on-the-side computation.

2. Punched Cards

The punched card system of transcribing data and instructions is extremely popular and convenient because it provides the opportunity to examine visually the information being processed for computer use. Magnetic tape, of course, usually does not provide this feature. A common procedure for high-speed computers is to use punched cards to transcribe data to magnetic tape. Thus the punched cards act as an intermediate input stage.

The format of the punching system is dictated by the requirements of the computer system. The standard punched card code, discussed earlier in this section may be used, but the transcription from punched cards to magnetic tape must contain format manipulation.

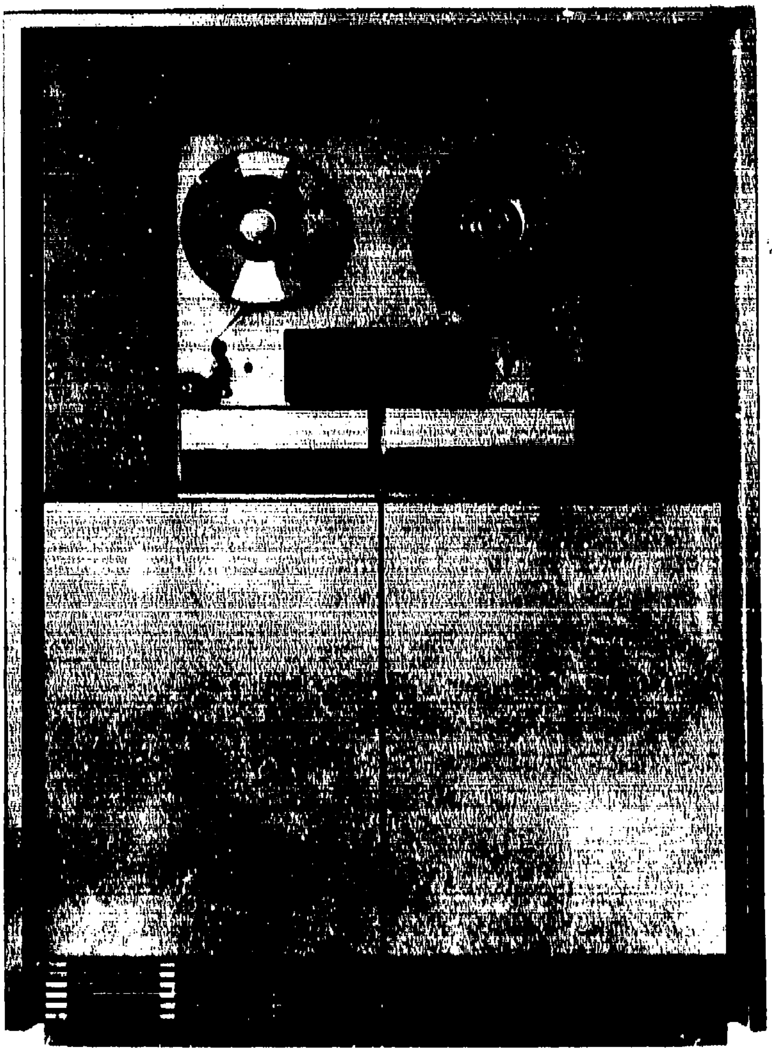
Punched card readers operate at speeds of about 600 cards per minute. The cards are read either by brushes which sense the punched holes by completing electrical contact through the holes, or by a photoelectric system sensing scheme. In either case, the data are composed of on-off signals, corresponding to the desired binary data form.

The card readers usually contain an error-sensing feature to check the accuracy of the punching or the reading of the card. The card may be read twice, and the two readings must not only agree, but must have correct parity (either an even or odd number of "1" bits).

If the punched card information is transcribed to magnetic tape, the

---

*Radio Corporation of America, Camden, N. J.



Radio Corporation of America, Camden 8, N.J.

Figure 120. Magnetic Tape Station

cards may be disposed of after the transcription, and the magnetic tape retained for future use of the data.

3. Punched Tape Reader

A punched tape reader uses a long strip of paper tape that is perforated in some type of digital code, similar to that used in the magnetic tape system. Either

five- or seven-track tape may be used, with both types having an additional track, consisting of a sprocket hole down the center of the tape to aid in guidance and to provide a timing pulse for the reader. The tape is read by a photoelectric system. One paper tape reader, the RCA 503, has entry speeds up to 1000 characters per second, which is fast enough to feed directly into the computer memory through the buffer input stages.*

Punched paper tape has the same advantage as punched cards: visual indication of data. The Flexowriter paper tape reader and puncher permits the accuracy of punched tape to be verified because it prepares a typewritten copy while it is punching tape; in addition, it can prepare typewritten copy from tape already punched.**

4. Analog-To-Digital Converters

For on-line operations of the digital computer in a physiological monitoring system, an analog-to-digital converter is connected to the input of the computer and information is fed directly to the computer or to intermediary magnetic tape storage. The information must be in a form that the computer can use, namely in a digital code format corresponding to the code and format for which the computer has been designed.

For many computers, binary-coded-decimal (BDC) is acceptable as an input language. In the binary-coded-decimal code (8-4-2-1 code), each decimal digit is represented by a four-bit binary code. Thus four lines per decimal digit are required in the computer input.

There are both electronic and mechanical methods used for the analog-to-digital conversion. The method used for the actual conversion differs according to the following criteria and requirements (ref. 6):

(1) Form of the analog data (time, frequency, voltage, shaft position, graph recordings, etc).

(2) Digital code output required.

(3) Number of bits in the output (determined by the precision desired).

(4) Quantization error. The error imposed by the choice of the number of bits.

*RCA Computer Model 503 Paper Tape Reader
**Flexowriter, Friden, Inc.

(5) Accuracy. Degradation caused by drift, miscalibration, etc.

(6) Holding characteristics. The necessity for providing a constant input to the converter during the period of conversion to serial binary form (may require use of a storage register).

(7) Conversion time. The time required to finish one coding conversion.

(8) Conversion rate. The number of conversions performed per unit time.

(9) Total or incremental determination. Whether the converter recalculates the digital value as an entirely new step, or determines the difference in value from the previous calculation and algebraically adds the difference in value to the value previously determined.

a. Electronic Analog-To-Digital Conversion

The following are four methods of electronic analog-to-digital conversion:

(1) Ramp Method. The ramp method, discussed previously in Section III under the heading "Digital Display," actually measures the time necessary for a linearly rising ramp voltage to reach the amplitude of the input voltage being converted. An oscillator is started when the ramp starts and it stops when the comparator circuit indicates equal amplitudes. The oscillator output drives a digital counter which has a coded output equal to the number of pulses received during the period corresponding to the value of the voltage digitally coded (see figure 56).

(2) Feedback or Comparison Method. In this method of analog-to-digital conversion, the digital output of the converter is compared to the analog voltage at its input. In order for the two quantities to be compared (digital and analog), it is necessary to convert the digital to an analog form, which is done with an internal digital-to-analog converter. Figure 121 illustrates this arrangement.

The output and input analog voltages are compared in the comparator circuit, where the output is subtracted from the input. If the input is greater than the output, a positive voltage is produced by the comparator. This positive voltage is applied to the digital decision switching unit, which causes the switching unit, a form of digital counter, to count upward to change the digital output to a higher value code. This new digital code is converted and compared and the process repeated, until the point where the comparator circuit produces a zero difference. This means that the output code is exactly equal to the input voltage and there has been an accurate analog-to-digital conversion. Of course, if the original digital

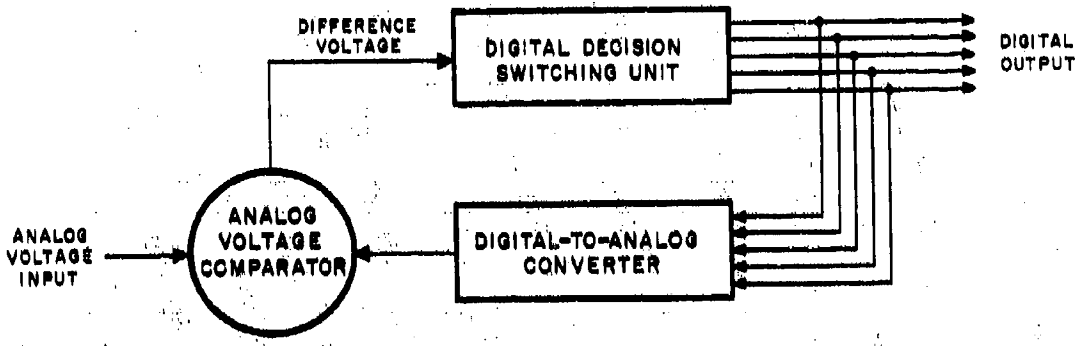


Figure 121. Feedback Method of Analog-To-Digital Conversion

code were greater than the analog voltage input, the subtraction process would have produced a negative difference, which would have caused the switching unit to count down and decrease the digital code output, until a zero difference was obtained.

The comparison and conversion steps in this type of converter must be performed very quickly, so that the output of the converter is able to follow changes in the input. The holding characteristic also is a determinant in the use of this converter, because the input must be of a constant value until all of the feedback trails are performed and the output has settled to the final value.

(3) Cascaded Stages of Comparators. A parallel cascading of comparators is used for analog-to-digital conversion, with each comparator set for a different level of input voltage corresponding to the value of the digital bit. The comparators are interlocked so that as one stage is turned on, the stage with the lower (or upper) value is turned off.

(4) Doubling Converter. The doubling converter does not require a digital-to-analog converter, as does the feedback method. In this system each digital bit is generated in sequence, from the most significant to the least significant. Figure 122 illustrates the operation of this converter (ref. 22).

The reference voltage is equal to one-half the full-scale conversion range of the converter. On the first sampling, the input voltage is compared to this standard voltage. If the input voltage is greater than or equal to the reference, the standard voltage is subtracted from the input voltage. At the same time the most significant bit has a "1" generated. After the subtraction, the voltage difference is doubled and again compared with the standard voltage. If the initial comparison indicated that the reference voltage was greater than the input voltage, then a "0" is generated for the most significant bit, no subtraction takes place, and the input voltage is doubled to be compared again with the reference voltage. This process is repeated over and over until all of the bits are generated. Thus the time necessary to perform a conversion is equal to the number of output bits times the conversion time for each bit.

To illustrate the operation of this converter, assume that the full-scale range of the device is 32 volts. The analog input voltage (unknown) is 27 volts. One-half of the full-scale range, 16 volts, is the value for the reference standard. On the first sampling, the input voltage (27 volts) is compared with the standard (16), the difference is positive (11 volts), and the most significant digital bit is generated as a 1. The difference voltage is doubled (22 volts) and then compared again to the reference. Again the reference is smaller (by 6 volts), the next most significant digit is generated as a 1, and the difference voltage is doubled. On the third sampling, the doubled voltage is only 12 volts, which is less than the reference voltage, so that the next bit generated is a 0. The voltage is doubled again without being subtracted (24 volts), subtracted, a 1 is generated, and the difference is 8 volts. When doubled and compared, the reference voltage is found to be exactly the value of this 16-volt level, so that a 1 bit is generated for the fifth place. Thus the digital output for the 27-volt input is 11011, which is 27 in binary notation.

b. Mechanical Analog-To-Digital Conversion

(1) Shaft Position Method. The shaft position method of another form of on-line analog-to-digital conversion involves changing the mechanical position of a code wheel attached to the rotating shaft. The code wheel has imprinted on it a pattern of conducting and nonconducting segments, which generate "1's" or "0's" according to the position of the segments in relation to brushes that ride on the segments (see figure 123).

(2) Graph-Recording Method. A second type of mechanical analog-to-digital conversion converts the results of experiments recorded on graph

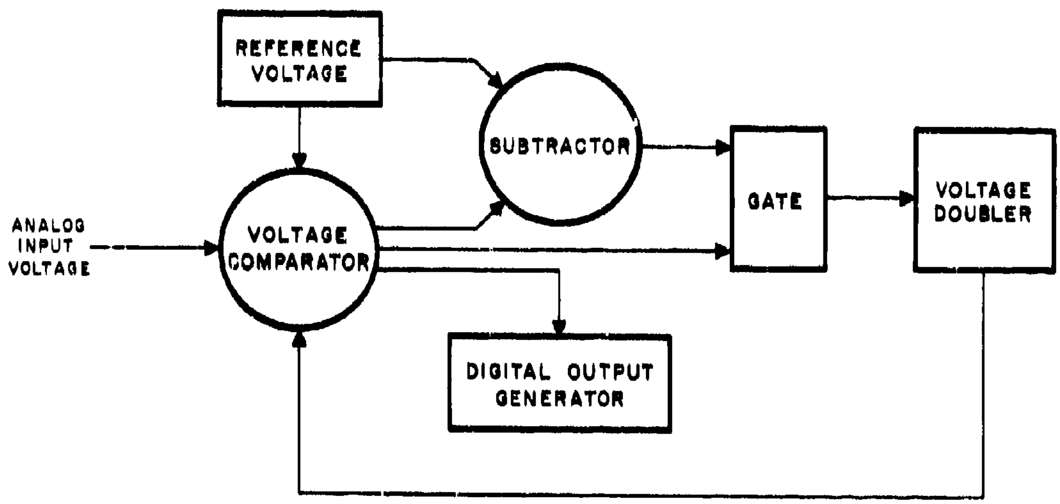


Figure 122. Doubling Method of Analog-To-Digital Conversion

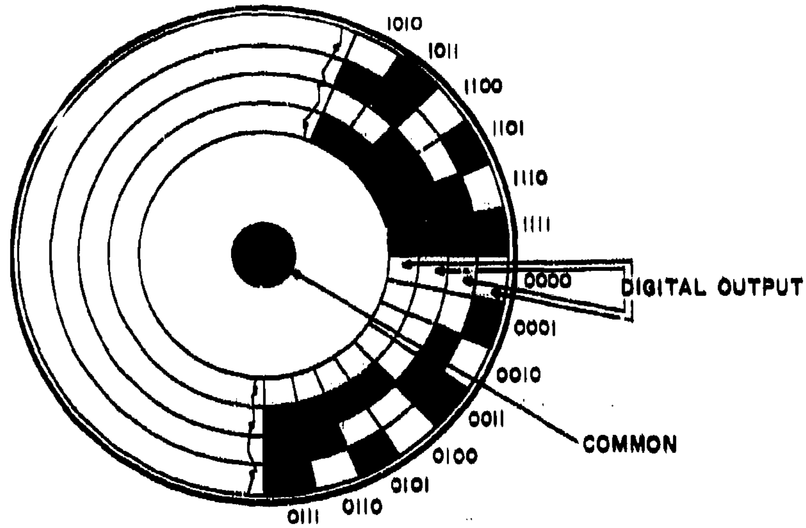


Figure 123. Code Wheel for Mechanical Analog-to-Digital Conversion

paper to a form acceptable to a digital computer. In one system* two crosshairs are aligned manually over the graphed function as the paper on which the data are recorded is transported at a constant velocity past the crosshairs. At uniform, preset time intervals, the position of the crosshairs is sensed, digitized, and readout in a form applicable either for direct input to the computer or for a paper-punched format.

A more automatic system is available, using the graph-recording method of analog-to-digital conversion.** In this system the graphed line is sensed automatically, either electromagnetically by means of a high-frequency signal passing through a silvered line drawn over the graphed line, or optically by sensing the pen line directly. The sensing head is connected to a follower servo system, which permits continuous tracking of the variation of the graphed curve, as well as providing output voltages corresponding to the graph amplitude. While the output is an analog voltage, this voltage may be digitized by one of the electronic analog-to-digital converters mentioned previously.

5. Digital-to-Digital Converters

When the input data to the computer are in a digital form but not in the format acceptable to the computer, a stage of digital-to-digital conversion must be used. In this converter, the following functions may be performed:

*Gerber Scientific Instrument Company, Hartford, Conn.
**F. L. Moseley Company, Pasadena, Calif.

(1) Change in number of bits per character.

(2) Change of bit polarity convention.

(3) Conversion from straight binary to binary-coded-decimal, or the reverse.

(4) Change of number of characters per word.

(5) Insertion or deletion of end-of-word symbol.

(6) Insertion or deletion of line characters (end-of-line or beginning-of-line) used with the typewriter transcriber.

(7) Change of amplitude level.

6. Typewriter

An on-line typewriter often is used with a computer system to enter small amounts of data manually. The typewriter is similar to a teletype device, where the output from the keyboard is converted to a binary code as well as printed copy for immediate perusal and error checking.

B. Output Devices

1. Magnetic Tape

Magnetic tape is used extensively as an output device for the same reason that it is used as an input to the computer: its fast speed of data transcription. The magnetic tape recorder operates similarly both as an output and input device, and in fact, the same machine may be used as both an input and output device during the same program. When the magnetic tape is being used as an output device medium, tape must be used that has been erased magnetically beforehand so that no false bits are left on the tape when the output of the computer is recorded. Most tape stations have a built-in erasing system so that the tape is erased before it reaches the recording head of the machine. Bulk tape erasers are used if the entire tape is to be erased at once, and they generally do a better and faster job than the built-in erasing circuits.

2. Punched Cards

While punched cards may be used as the on-line output medium of the computer, their relatively slow punching speeds (a few hundred cards per minute) restrict their use in most computer systems. Instead, they often are used as the secondary output device, being controlled by a magnetic tape station which has had the original computer output data stored in it.

3. High-Speed Printer

High-speed printers often are used as output devices for the computer because their writing speeds approach the speed of the computer. One such high-speed printer has the capability of typing 600 lines a minute, with each line having a maximum of 120 characters. This rate of 1200 characters per second is practical for most on-line applications.

Two types of high-speed printer mechanisms are used. In one model the letters to be printed are contained on a print wheel which rotates at a relatively high speed (900 rpm). There are 120 faces on the print wheel, with each face containing all the letters, numerals, and symbols used in the system. As the information from the computer output is entered into the high-speed printer, the angular position of the print wheel is coordinated with the symbol to be printed. As the wheel turns, solenoid hammers are actuated which strike the paper onto the print wheel through a ribbon. Fast printing speeds are possible because the only mass to be moved is the relatively low hammer mass. The more massive type and carriage are not repositioned for each letter, as in a standard typewriter.

The second method generally used is the wire matrix printing technique. In this type of printer each character is formed by a combination of small wires in a 5 by 7 wire matrix. Figure 124 shows this type of configuration. The tips of the wires are inked or press an inked ribbon onto the paper in the printer. Because of the extremely small mass of these wires, very little inertia is associated with the mechanical portion of the printer, allowing the high printing speeds necessary for on-line applications. The characters printed by the wire-matrix method, although not as legible as with the print wheel, are adequately readable.

Other methods such as optical and photographic techniques forming the print-out information are used; however, for most general purpose computer applications, the additional complexity and cost are not justified to obtain the higher printing speeds achievable.

4. Typewriter

The same typewriter used to enter data in small amounts into the computer may be used as an output device if the amount of data is limited to a few numbers or lines of numbers. The typing speed of the typerwiter is only about 6 to 12 characters per second. The typewriter also can be programmed to give a plot of the output data by arranging the computer output format in a character-versus-line pattern. Either output can also be displayed on a cathode-ray tube, and photographed.

5. Punched Paper Tape

Punched paper tape may be used as an output medium as well as an input

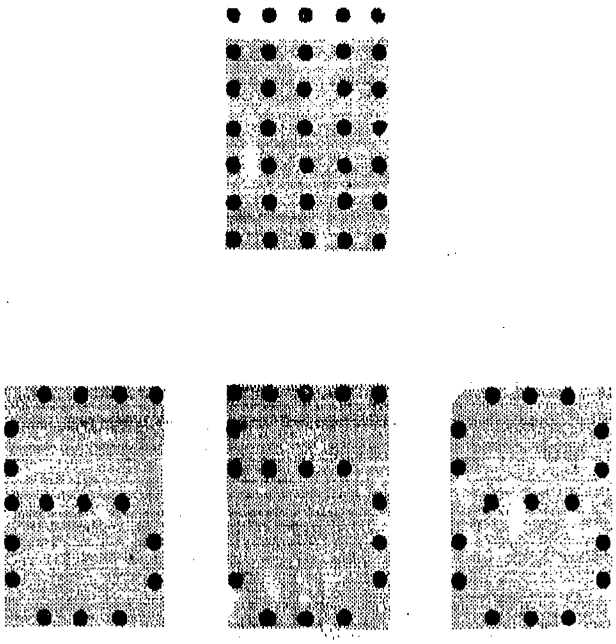


Figure 124. Wire Matrix for High Speed Printer

medium. The speed of the punch is relatively slow, ranging from 20 to about 300 characters per second. Either 5 or 7 punch holes per character formats are available.

C. Computer Registers

In the arithmetic unit of the computer are elementary memory or storage devices called registers. A register is a circuit with the capability of storing the bits of information comprising a computer word. When the registers are connected so as to add (in binary) the arithmetic contents of one register to the contents of another, the combination of the register and the adder circuitry is often referred to as the accumulator. Frequently, three registers and an adder compose the arithmetic unit: one register stores the addend, the second register stores the augend, and the third register stores the sum. In multiplication and division, these registers would contain, respectively, the multiplier or quotient, the multiplicand or divisor, and the product or dividend.

Certain registers also have the ability to shift the information stored in them one unit right or left. This feature may simplify multiplication and division. When entering data from a serial source into the register, a shift register is used to clock in the data one bit at a time into an empty register until the full word length is stored; thus serial-to-parallel conversion is performed. Also, normalization and round-off of numbers by shifting the number right and then left in the register, eliminating the least significant bit.

The control unit of the computer also has registers. The registers here are used to store temporarily the addresses of the memory locations for the data that is to be sent to the arithmetic unit.

The input data to memory locations and the output data from the computer memory also go through a register, called a buffer register. This register compensates for the difference in the data-handling speeds of the input/output devices and the computer memory.

The registers are composed of series of storage circuits called flip-flops. These circuits are so named because for each input pulse (bit), they change their state from an on-to-off or from an off-to-on condition. In a parallel register each flip-flop stores one bit. The flip-flop circuits in most modern computers are transistorized, with a minimum of two transistors per flip-flop necessary.

The computer registers, as mentioned previously, are used only for temporary storage. These register storage areas are designed specifically to allow the arithmetic operations to be performed with the greatest degree of efficiency, accuracy, and speed. By separating the functions of the storage and arithmetic areas, the computer size and cost are reduced. The more specific memory elements (described below) are used to store the large quantities of data necessary for complex calculations; the elementary operations are performed in the computer registers.

D. Memory Devices

One of the characteristic components of the digital computer is the built-in storage or memory section. If the computer did not have this memory capability, the computer operator would have to feed into the computer each element of information singly as it was needed, and the computer would be just an oversized and very expensive desk calculator. With the memory capability of the computer, however, the operator can feed instructions and input data to the computer, and the computer proceeds automatically, processing the data according to the instructions, assembling the desired answers, and making the results available at the output in whatever form the instructions dictated. Certain routines that are used frequently, but not in every computation, may be stored externally (for example, on a reel of magnetic tape), and made available to the computer when requested.

Computer memory devices are classified in different ways: according to the organization of information in the memory; whether the withdrawal of stored information erases the memory of this information permanently; and according to the speed of withdrawal and insertion of information out of and into the memory. The following is a discussion of the major memory devices.

1. Magnetic Tape

Magnetic tape, often used as a memory device in a digital computer, is seldom used exclusively, but rather as an auxiliary storage capacity associated with some other memory system. Magnetic tape is considered as a low-speed, high-quantity form of storage with a nondestructive readout capability. Since magnetic tape is wound upon large reels, information required by the computer may be stored at one end of the tape when the reel is positioned with the other end of the tape at the reproduction head. The tape would have to run its full length (requiring a minute or more) before the information became available. When information is read from the tape, the tape is not disturbed, since the reproduction head merely senses the magnetic state of the tape.

As mentioned previously, tape is an excellent computer input and output device. It also may be used to handle the overflow of information from the high-speed memory of the computer, and as a storage means for certain routines retained in the computer facility library for reuse. These routines usually are read out of the tape into the high-speed memory of the computer when needed.

2. Magnetic Core

Magnetic core memory is the most widely used form of high-speed computer memory storage because of the extremely high speed of information retrieval and high density of packaging possible despite its relatively high cost. The cores are small ferromagnetic doughnuts (figure 125) with wires threaded through the holes. The magnetic operation of these cores is such that only two stable states of magnetization exist. These two states are reached by causing a current of sufficient amplitude to flow through one of the wires threaded through the hole of the core or, alternately, to have the sum of currents in more than one wire through the center of the core great enough to overcome the coercive force of the magnetic material. The choice of the two magnetic states is determined by the direction of current flow through the wires.

Figure 125 illustrates the operation of the magnetic core memory.* Three wires pass through each core: the write winding, which carries a current of 100 milliamperes when energized; the read winding, which is looped twice around the core, so for a read current of 210 milliamperes, an effective magnetization force of 420 milliampere-turns is present; and the sense-digit winding, which for a write cycle is energized with 100 milliamperes in the same sense as the write current, and in the read cycle is a pickup lead for induced voltages. The magnetic core is in one of the two states indicated on the figure: the "0" state corresponding to positive B and H, and the "1" state corresponding to negative values of B and H. The core switches from one

*Ampex Model 8192-LQ-56 High Speed Memory, Ampex Computer Products Company, Culver City, Calif.

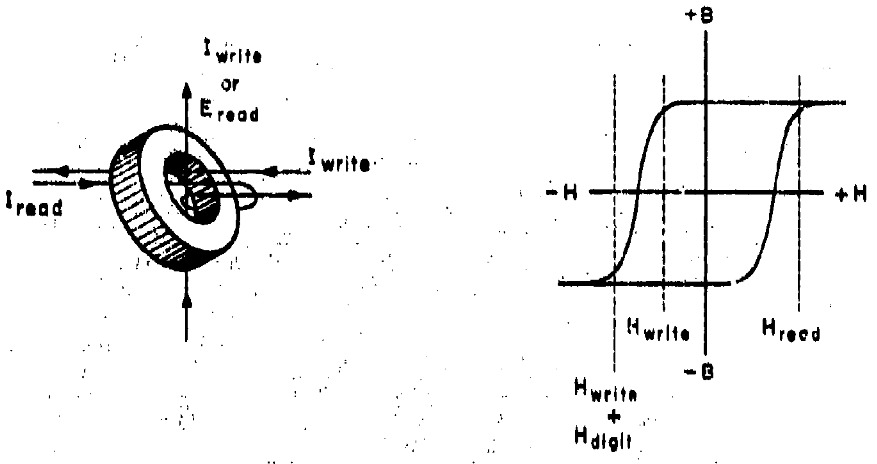


Figure 125. Magnetic Core Memory Operation

state to the other very quickly if the H is reversed quickly. The H is produced by the current through the wires around the core. Approximately 200 milliampere-turns are required to switch the core from one state to the other. Thus, if the core is in the "0" state and both the write and sense digit windings are energized, it will switch to the "1" state; if the core is in the "1" state and the read winding is energized, the core will switch to the "0" state. When the magnetic state is changed, a voltage is induced by the change in flux. Therefore, if the core is in the "1" state and the read winding is energized, a voltage is induced in the sense winding. If the core is already in the "0" state, no voltage is induced. The magnetic state of the core therefore can be sensed and provides a convenient method of memory readout. For each readout, the core is returned to "0"; if the information the core is storing must be retained, a "1" must be regenerated to reset the core.

The cores are strung in a wiring matrix in such a way that each write line threads through all of the cores associated with certain 56-bit words (see figure 126). At one end of each write line is a write-drive switch; a sink switch is at the other end. By properly activating one write switch and one sink switch, only one set of cores of the 8192 word locations in the high speed memory has the write current passing through it. However, the write current alone is not great enough to switch the state of the cores.

The sense-digit lines each go through one bit (one core) of each of the 8192 words. When a sense-digit line and the write lines are both activated, only one of the 458,752 cores (8192 words x 56 bits per word) in the memory has the full 200 milliamperes necessary to switch its state. Thus a "1" may be written into a given core location, corresponding to one bit of a given word. Actually, all 56 bits of a given word are written at the same time in this parallel operation memory.

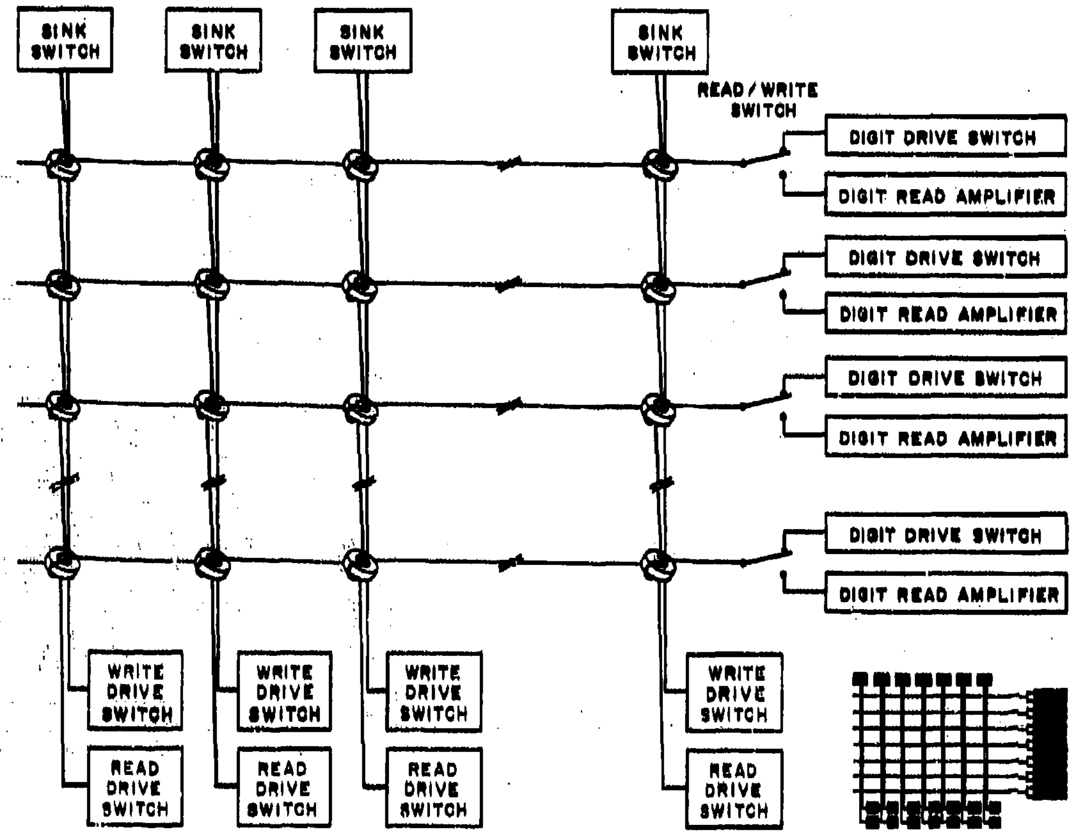


Figure 126. Magnetic Core Memory Switching Circuit

The readout of information from the core memory is the same as the writein, except that read-drive switches are activated in conjunction with the sink switches. The sense-digit lines are connected to amplifiers to sense the induced voltages, as contrasted to their connection to digit drive switches during the write operation. Figure 126 illustrates an array of cores connected to the read and write switches.

The typical operation time of a magnetic core memory for writing or reading is about 5 microseconds. This type of memory is classified as a destructive memory for the information is destroyed upon readout. However, if a loss of power to the memory should occur, the magnetic cores retain the stored information.

3. Magnetic Drum

Magnetic drum memory often is used as an auxiliary memory or as the

main memory in smaller computers. The memory device is a large rotating cylinder whose surface is coated with a magnetic material similar to the coating on magnetic tape. A number of heads are positioned in a line parallel to the long axis of the cylinder, forming tracks of information. Both read, write, and erase heads are positioned over each track.

As the drum rotates, the write head writes bits of data on a track. For the retrieval of information, the computer must wait until the storage location on the track is under the read head, therefore, the access time for magnetic drum memory is not fixed, but depends on the drum location and the rotation rate. An average access time of 10 milliseconds is required when a magnetic drum is used.

To make the drum at all efficient, parallel readout is used. Thus, in the 10 milliseconds necessary to retrieve information, the whole computer word is retrieved. In serial construction an average of 10 milliseconds is required to reach the first bit of the word, and the remaining time needed for the additional bits in the word is a function of the rotation speed of the drum.

As in magnetic tape applications, the magnetic drum memory is not erased upon readout. The information, of course, can be erased and new information entered upon command of the computer. Often the drums have one or more permanently recorded tracks which supply a timing pulse for the computer as well as a reference signal indicating where in the drum rotation cycle the drum is at any instant of time.

4. Magnetic Disc

The magnetic disc memory is used only as an auxiliary memory because of its extremely slow access time. The device is similar to an automatic juke box with discs of magnetic material instead of audio records. On both surfaces of each disc are recorded the bits composing the computer words. One such disc file memory* has 128 discs, allowing a storage capacity of over 4.6 million characters. By having such a large capacity computer memory available (even with the relatively slow access time memory), the computer is able to process tremendous amounts of data without exhausting all storage space. The various service and translating routines also may be stored without using limited, high-speed memory room.

The access time for such a disc file is about 2 seconds. The actual data transfer rate between the high-speed memory and this external low-speed memory is 2500 characters per second. As in the other magnetic writing systems (tape and drum), the information may be read without destruction or it may be erased and recorded at will.

---

*RCA 301 Disc File, Radio Corporation of America, Camden, N. J.

5. Acoustic Delay Lines

In some of the early computers (EDVAC, UNIVAC 1), an acoustic delay line was used for high-speed storage. In this system a long tube of mercury has two transducers inserted at each end: one a transmitter and the other a receiver. A burst of ultrasonic energy is introduced at one end, and this pulse travels through the tube at the speed of sound in mercury (1.45×10^5 cm per second), and a delay of known magnitude is introduced. At the end of the tube, the pressure wave is picked up, amplified, and reintroduced to the transmitter. This recirculating memory was found to be rather inefficient because of its long access time (100 microseconds), its volatility, and its large size. Magnetostrictive delay lines using nickel alloy wire are also used.

6. Other Memory Devices

Some additional memory schemes have been developed recently, including the magnetic film memory (similar to the magnetic core), diode-capacitive memory, and electrostatic tube memory. Both the magnetic film and the diode-capacitive methods promise high density and fast access time.

III. Operation

A. Programming

The digital computer cannot think. It knows nothing more than what it is told and it will do nothing more than what it is instructed. It cannot anticipate and it cannot surmise. It cannot use partial information to achieve a complete solution, except by statistical guessing, and the method used in the guessing must be spelled out in minute detail. As a result of these limitations, one of the major areas of interest in automatic data processing is the conveying of information and techniques from the human language to that of the computer. This is the task of the programmer.

1. Machine Language Code

The program prepared by the programmer for the computer must be intelligible to the computer. As discussed before, the computer basically is sensitive only to binary language: the language of off-on, 1-0, yes-no. Either the programmer must be able to write in binary or else a translation program must precede the actual computation process of the computer. Computer programs often may be composed of symbology based on the alphabet and the decimal system, which the computer interprets. Other machines require an all numeric input, while others require that the input be in binary form from the start.

The program written for the computer consists of statements (data statements such as A = 1) and commands (equations like $D = B^2 - 4AC$; find D). Without

previously instructing the computer how to handle a command written out as a complete equation, the command must be broken down into simple distinct steps, with each step employing one arithmetic operation (addition, multiplication, etc). As well as being able to follow these simple commands in order, one-by-one, the computer can also be instructed to jump from one set of steps to another, depending on the results obtained after the first set of instructions. Thus the computer can examine an answer it has calculated, compare the answer with some other quantity stored in the computer memory, and modify its next step if the calculated answer is greater than the stored number. By writing the computer program in such a manner, the computer can repeatedly perform an operation over and over again until the desired number of operations has been done, or until a desired accuracy has been obtained.

To program a computer to perform any arithmetical process, a knowledge of the computer code is necessary. On a large scale computer, about 50 operations are wired permanently in the computer as part of the normal functions the computer can perform. These instructions can be broken down into four categories: input/output, arithmetic, data handling, and decision and control.

The input/output instructions provide for data entering the computer by way of one of the input devices and leaving the computer to be received by the output devices. These instructions are concerned with tape movement and readin/readout, typewriter printin/printout, on-line printer printout and paper advance, and punched paper tape readin/readout.

The arithmetic instructions include addition, subtraction, multiplication, division, and comparison for equality (a subtraction process). Some machines have provisions for performing these arithmetic operations by either decimal or binary methods.

Data-handling instructions include the transfer of information into, out of, and between locations of the high speed memory, the shift right and left in memory locations, and the search for specific data in input information located on magnetic tape or punched tape.

The decision and control instructions include stop and start instructions, the transfer of programs for subroutine entry, and the temporary transfer to another program or to input/output tasks in the middle of a computer computation cycle.

Each computer has a definite format for the writing of an instruction, which depends on the internal logic of the computer. An instruction consists of two parts: the operation, and the memory address or location of the data on which the operation is performed or stored. Some machines require more than one address, specifying not only where the data are stored, but where the answer to the computation is to be sent. Machines therefore are classified as one-, two-, three-, or four-address instruction format computers.

In a four-address format, one address specifies one operand in the computation (the addend, minuend, multiplicand, or dividend), the next address specifies the second operand (the augend, subtrahend, multiplier, or divisor), the third address instructs where the result of the computation should be stored, and the fourth address informs the machine which instruction to perform next. Thus an instruction might be of the form shown on figure 127(A).

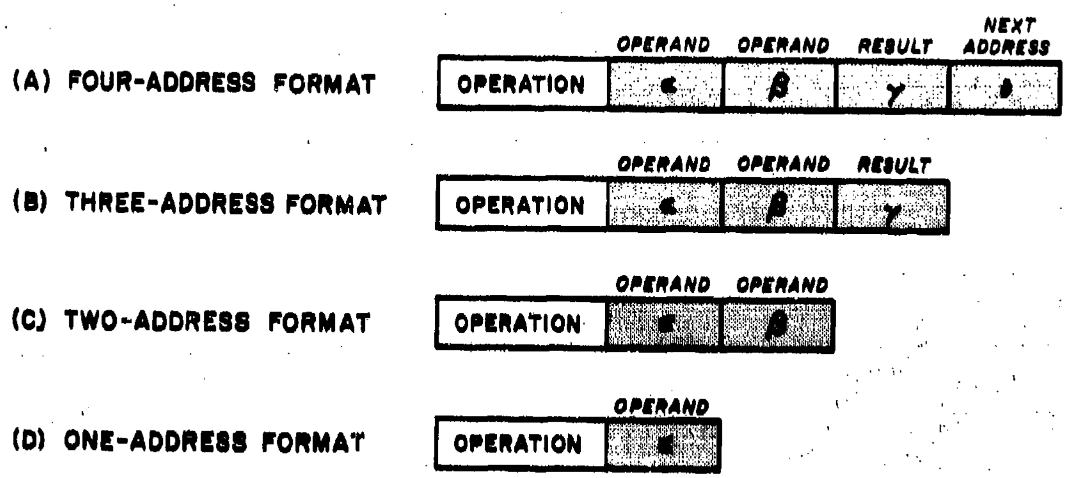


Figure 127. Computer Instruction Address Formats

The fourth address is eliminated by having an address counter in the computer and specifying that the next instruction will always be located in the following memory location; e.g., if the first instruction is stored in memory location 00001, the next instruction will be found in 00002. (See figure 127(B).) With this arrangement, a special conditional jump instruction must be included if sequential operation is not desired (as in an iterative numerical calculation).

An instruction is reduced to two addresses by eliminating the third address, the location where the result is to be sent. Instead, a permanent result register, called the accumulator, is provided, and all results are automatically put into the accumulator. (See figure 127(C).) An additional instruction transferring the contents of the accumulator to a memory location is required in the two-address instruction format.

In the one-address instruction, the accumulator is used not only for the result, but also for the second operand. For example, in a problem in addition, the augend is put in the accumulator (a transfer instruction), the addend is added to the accumulator and the sum put in the accumulator, and then the sum is transferred from the accumulator to the desired memory location.

The number of addresses specified in the instruction determines the number of instruction steps necessary to perform a given operation. The fewer addresses that are specified, the more steps that are necessary; with an abbreviated instruction, however, smaller computer words are used, requiring a smaller number of bits per memory locations, and allowing more word storage for a given memory size. The ultimate choice depends on the type of problems expected to be encountered and the relative cost of memory and accumulator hardware.

Figure 128 illustrates a sample program for a hypothetical computer. A three-address machine is used and a simple calculation involving subtraction and multiplication is performed. The memory locations are specified in decimal notation, although octal notation is usual in computer programming. Each memory location can store an 8 (decimal) digital number.

Twelve memory locations (00 through 11) are used to perform the operation of forming $B^2 - 4AC$. (Less could be used since both B and 4 have the same value and one memory location for the storage of "4" would suffice.) The necessary operation instructions are loaded into locations 00 through 05; 06 and 07 are assigned to carry results, both interim and final; and 08 through 11 are loaded with the constant values involved in the computation. After the computer is started (manually) with the instruction in location address 00, the remaining operations take place, in sequence, automatically. The content of memory location 08 (A) is multiplied by the content of memory location 09 (C). The product is put into location 07, which was deliberately left empty for this purpose. The product AC in 07 then is multiplied by the content of location 10 (the constant 4). This product, 4AC, is put back into location 07. (When a new number is placed into a memory location, any previous number is erased, leaving a clear location for the new data.) The quantity B is squared by multiplying the content of memory location 11 by itself, and the product is put in empty location 06. The content of location 07 is subtracted from the contents of 06, providing the desired quantity, which is stored in location 07. The computer then stops. The content of location 07 (the answer of the calculation) can be printed out with an appropriate instruction.

It may be noticed that memory locations 00 to 05 contain instructions, while locations 08 to 11 contain input data. The data statements are treated as a single 8 digit number (thus location 09 contents is 00000002) with no significance associated with the operation, α , β , or γ column. However, as far as their general form is concerned, the two types of notation are identical, both consisting of 8 decimal digits. Extreme care is necessary in the writing and transcribing of a computer program, because, if by accident an instruction were treated as a data element or a data element treated as an instruction, the computer could not detect the difference, and the machine would either provide an erroneous result, go into a repeating cycle, or stop. In any case, the program would have to be examined carefully to determine where the mistake was made, a task which can take many hours or days, depending upon the complexity of the problem.

Problem: Find $D = B^2 - 4AC$.

$B = 4$
$A = 1$
$C = 2$

INSTRUCTION TABLE

Code	Operation	Meaning
00	START	Go to next instruction.
01	ADD*	Add (α) to (β); put sum in γ .**
02	SUBTRACT	Subtract (α) from (β); put difference in γ .
03	MULTIPLY	Multiply (α) by (β); put product in γ .
04	STOP	Stop computer.

PROGRAM

Location Address	Instruction: Operation Code	Instruction: α	Instruction: β	Instruction: γ	Remarks
00	00	00	00	00	Start (MANUAL).
01	03	08	09	07	Multiply A by C.
02	03	07	10	07	Multiply AC by 4.
03	03	11	11	06	Square B.
04	02	07	06	07	Form D.
05	04	00	00	00	Stop (AUTOMATIC).
06	00	00	00	00	
07	00	00	00	00	
08	00	00	00	01	Constant A.
09	00	00	00	02	Constant C.
10	00	00	00	04	Constant 4.
11	00	00	00	04	Constant B.

*The ADD operation is not used in this sample problem.
**(α) means the contents of address α .

Figure 128. Sample Program for a Hypothetical Computer

2. Algorithm Computer Code

The previous discussion concerned a computer program written in a form relatively basic to the computer. Each instruction was in digital form, the addresses for the computer were specified, and the actual step-by-step program was detailed. For this type of programming to be effective, the programmer must know the computer construction so that he may assign the proper memory locations, perform the logical steps for which the computer was designed to perform, and effect readin and readout of data and results. If all programmers needed this knowledge, the number of programmers would be limited severely. To make the computer facility available to the scientist who is not a full-time programmer, but wants to use the computer on a frequent schedule, a special method of computer programming has been devised.

Certain computer programs, known as algorithm statement programs, have been formulated to decrease the difficulty involved in computer programming. In these programs, word abbreviations or symbols familiar from arithmetic notation are used in place of the abstract digital commands which the computer readily understands. Instead of using the symbol of 01 to indicate an addition process (as was specified in the sample program), the word ADD or the symbol + is used with the appropriate letters standing directly for the quantities to be added as opposed to the memory locations where the data are stored. The FORTRAN programs (FORmula TRANslation) prepared for the RCA computers use words and symbols like +, SQRTF (square root), and parentheses to set apart each quantity. Thus, an addition and square root statement would read:

SQRTF (A + B)

A typical FORTRAN program is shown in figure 129.

International standardization of a symbolic program code has been attempted, with ALGOL (ALGOrithmic Language) evolving. In this symbology, most of the arithmetic steps use symbols, such as +, -, X, and /. Thus a statement of the sample problem shown in figure 129 would be:

;A:=1;B:=4;C:=2;D:=(+B↑2↓-4XAXC);stop;

The semicolons and colons are used to separate the factors in the program statement. The arrows indicate that 2 is the exponent of B. This is the entire program, and the computer then would read in the information, make appropriate memory assignments automatically, perform the calculation, and stop. Once a given program is written for one computer, the program is usable for all computers and thus duplicate programming is eliminated. Of course, a translating compiler program must be written for each manufacturer's computer that would translate the ALGOL statements into the digital statements understandable to the computer. In addition, this compiler routine automatically sets up the memory assignments, effects all necessary transfers, and arranges input/output formats.

```
*JOB        SAMPLE FORTRAN PROBLEM
*EXECUTE
*COMPILE
            READ 01, A, B, C
   Ø1       FORMAT (3E5.0)
            D = B**2 - 4*A*C
            PRINT 02, D
   Ø2       FORMAT (1E15.5)
            STOP
*DATA
   1.000    4.000    2.000
*END FILE
```

Note: The format identification (e.g., 3E5.0) defines the numbers that must be loaded in or printed out of the computer -- how many numbers, how many digits in each, and where to locate the decimal.

Figure 129. Sample FORTRAN Program

Presently, most manufacturers issue scientific program compilers with their computers. In addition, another type of compiler translator, incorporating arithmetic and processing functions more unique to the business use of the computer, has been written. This compiler program, COBOL (COmmon Business Oriented Language), was designed under Department of Defense auspices as the common language for data-processing by computers. It functions in a manner similar to ALGOL and FORTRAN.

B. Computer Operation

The steps necessary for operating the computer, once the program has been written, depend largely on the type of computer and the peripheral equipment available at the facility. The data and the program must be transcribed into a usable computer form. Punched cards often are used, as is punched paper tape. Some facilities

have provisions for entering the information directly onto magnetic tape by means of a magnetic tape writer*. If a punched card system is used, the contents of the cards are either transferred to magnetic tape or used as the input medium itself. (Magnetic tape of punched tape may be used directly.) The tape is mounted upon a tape station or tape reader, or the punched card reader is loaded with the program and data cards. The computer is set up to receive the input data from the appropriate input device and then is started.

The data are read into the memory of the computer (this phase is called loading the memory), the first program instruction enters the control register, and the computer starts computing. Depending on the program, the results of the computation are either stored in the memory for readout all at one time or the results are read out as they are obtained. The readout is either on magnetic tape, on a high-speed on-line printer, or on a paper tape puncher.

If the output device is magnetic tape, the data then are transcribed, at some later convenient time, by a card transcriber or an off-line high-speed printer. The punched tape also might be transcribed by a high-speed printer.

If a compiler translation program is used, or if any subroutine is called for in the computer program, the program, stored on magnetic tape or punched cards, must be obtained from the library of programs at the facility and be mounted upon a tape station or put in a card reader, so that it is available when called for by the computer.

ANALOG COMPUTERS

1. Components

The elements composing an electronic analog computer are the operational amplifiers (which are connected together to do the computing), the input devices (which introduce the constants and variables for the computation), the output devices (which display or use the computed results), and special devices (which perform nonlinear computations).

Basic analog computations may be performed without operational amplifiers or special devices. Simple analog addition is possible with a resistance network. Figure 130 shows such a circuit where the voltage across the output resistor R_4 is the linear sum of the voltages across R_1, R_2, and R_3. Similarly, simple integrating networks, discussed previously in Section II, may be composed of a resistance-capacitance circuit. The charge and thereby the voltage across C_1 (shown in figure 130) is the time integral of the charge or voltage supplied to the capacitor through R_5. These simple

*Remington Rand UNITYPER, for example

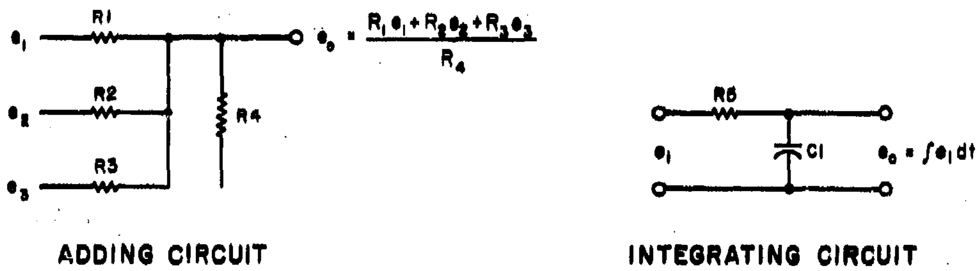


Figure 130. Basic Analog Computing Circuits

circuits do not suffice for accurate computations because, without amplification, the losses introduced by the imperfect resistance and capacitance components result in inaccuracies.

A. Operational Amplifiers

The operational amplifier (discussed in Section II) is the nucleus of the analog computer; it can integrate, add, scale, multiply by a constant, and invert. It must be a carefully designed, highly stable electronic circuit, because the computer is only as accurate as the amplifiers it employs.

The operational amplifier is a very high gain d-c bipolarity output device with exceptionally low-drift and wide frequency response characteristics. Open loop gains of 10^8 are typical. The bandwidth of one commercial device* is 1000 kc at the half-power points. Phase shift between the input and output is only 0.05° at 100 cycles per second.

The outputs for operational amplifiers range from ±50 volts dc to ±100 volts dc. For a positive d-c voltage at the input, a -100-volt maximum output may be obtained, while a negative d-c input provides a positive output voltage. Thus, an operational amplifier is basically an inverting amplifier. The actual gain of the amplifier is not 10^8. This figure is the open loop gain, which is the gain achievable with no output load and no feedback connection in the amplifier circuit. However, there is output loading and, most important to the operation of the amplifier as a computing device, there is a feedback connection.

Figure 131 shows the three connections used with the operational amplifier. The first connection is a straight amplifier connection, with the amplification factor equal to the ratio of the feedback resistor to the input resistor. This connection may be used to multiply the input voltage by a constant (the feedback-to-input-resistance

---

*Donner 3200 series amplifier, Systron-Donner Corporation, Concord, California

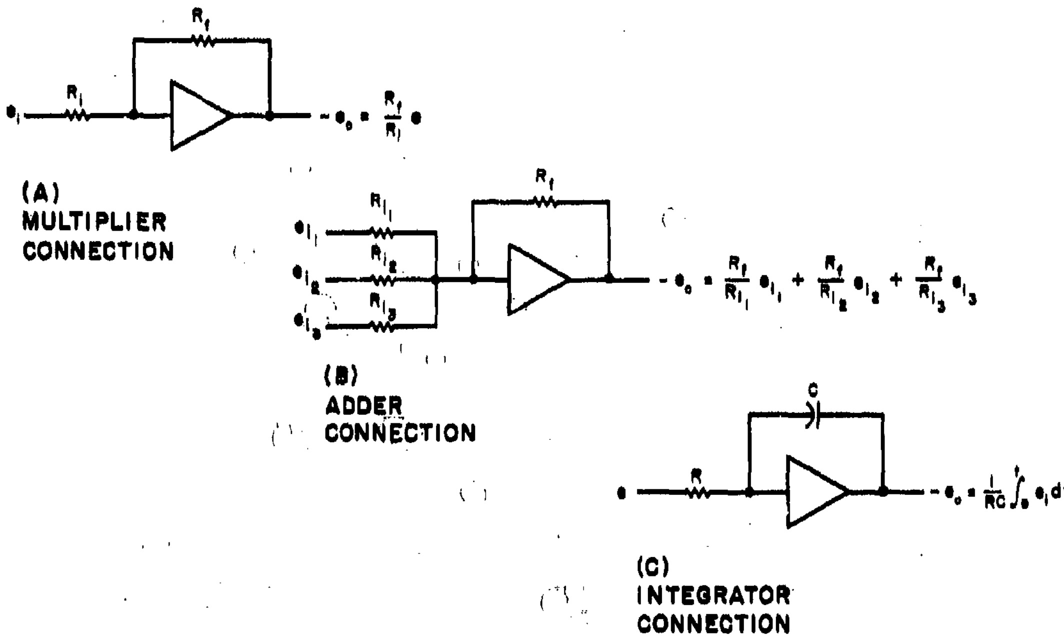


Figure 131. Operational Amplifier Connections for Analog Computer Operations

ratio). If the two resistors are equal, the amplifier performs a unit inversion. If the feedback resistor is smaller than the input resistor, the input is divided by a constant.

If more than one input is used, as shown in figure 131, then each input is multiplied by the ratio of the feedback resistor to the individual input resistor, and the output of the amplifier is the sum of the inputs. Generally, there is provision for no more than five inputs per amplifier. (Two or more adders may be connected in tandem for more inputs.)

The third connection shown is an integrating operational amplifier configuration. The feedback impedance is a capacitor, and the input impedance a resistor. It can be shown that the ratio of the feedback impedance ($1/2\pi fC$) to the input impedance (R) represents an integration constant multiplying the input voltage. The resistor-capacitor combination is a time constant circuit of value RC. The integration is being performed with respect to time, so that the expression for the output voltage is:

$$e_o = K \int_0^t e_i dt$$

(where $K = 1/RC$) or, expressed in another way, the output voltage is the time integral of the input voltage. By changing the resistance-capacitance product, the time factor is changed and thus the time during which accurate integration occurs may be varied.

By interconnecting these various configurations of operational amplifiers and including various scaling potentiometers, a problem to be computed is said to be mechanized. Basically the analog computer is the logical and systematic interconnection of these amplifiers in conjunction with various input and output devices so as to simulate the behavior of the real system under study.

B. Input Devices

The analog computer may be used to modify or compute mathematical functions of experimental or design data. In on-line applications, where the computation is being performed at the instant it is received from the monitoring channel, the input data may be the output of a telemetering system. Off-line applications use recorded data from experimental runs.

1. On-Line Information Devices

The data received from on-line monitoring applications must be in a form compatible with the input requirements of the computer. These compatibility requirements include: (1) voltage amplitudes of less than 100 volts, (2) frequency variations within the response range of the computer, and (3) freedom from noise and a-c pickup. Attenuation of the signal can correct excessive amplitudes. If the frequencies of interest are below 10 cycles per second, filtering can remove high-frequency variations in the signal. Good grounding procedures often eliminate excessive noise and hum in the input level. No particular input equipment is associated with on-line connections, except for an impedance-matching network that may be required to match the input line to the computer input terminals.

2. Recorded Information Transcription Devices

Recorded information may be available either on paper strip recorders or on magnetic tape. If the data are on magnetic tape, the output of the tape playback must have the same characteristics as for on-line computation. If the data are contained on a paper graph, some means must be used to transform the paper record to a voltage output, so that the information may be used with the computer. Transcribing devices that follow variations in graphical information, described under "Input Devices for Digital Computers" in this section, allow this transformation to be performed. However, the outputs of these devices are not digitized; instead, the continuous voltage is used as the analog computer input. The same restrictions listed above apply.

3. Digital-to-Analog Converters

If the information input to the analog computer is obtained from a digital system or from a telemetry system, the information often is in a digitally coded form, and a digital-to-analog conversion step must precede the entry of the information into the computer.

There are different methods of digital-to-analog conversion; the choice of methods depends upon the accuracy and speed of conversion needed and the cost. The following are three methods, two of which are somewhat alike. The amount of equipment involved in the second and third methods is small compared to the first.

a. Feedback Method of Conversion

The feedback method of conversion is the reverse of the analog-to-digital method described in the discussion of digital computers. As shown in figure 132, the digital input data are compared with the output of a voltage generator through an analog-to-digital converter. If the two inputs are identical in the comparison circuit, the output of the voltage generator is equal to the digital value of the input data. A difference in the two voltages results in an error voltage output from the comparison circuit, which, in turn, varies the output of the voltage generator to minimize the error signal.

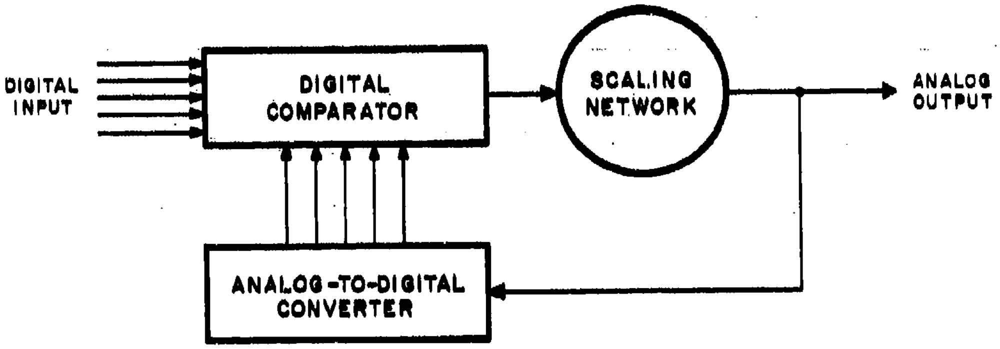


Figure 132. Feedback Method of Digital-to-Analog Conversion

b. Constant Current Summation Method

In this method, the input data are used to turn on or off a series of constant current output flip-flops (one flip-flop per input bit), as shown in figure 133(A). The values of the resistance network between the flip-flop outputs and the converter output are chosen so that the current contributed by each "on" flip-flop will be such as to make the output voltage proportional to the value of the digital number at the input.

c. Constant Voltage Source Conversion

This method is similar to the constant current model: the values of the resistors in the output circuit are selected according to the digital worth each switch possesses (see figure 133(B)). The output voltage is the sum of the individual voltage drops across each resistor, and, therefore, the analog output voltage is equivalent in value to the digital number at the input of the converter.

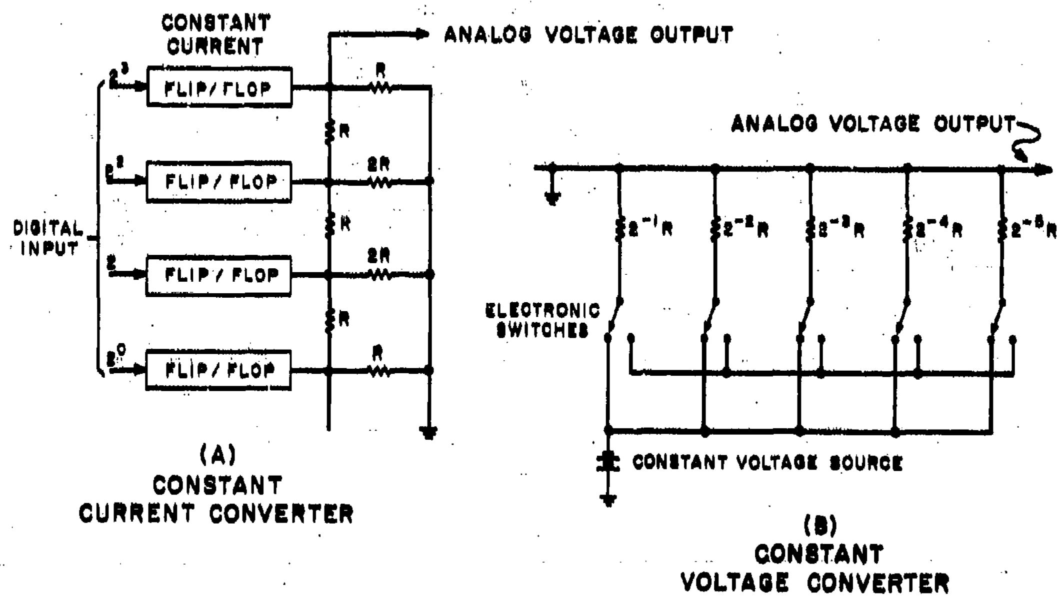


Figure 133. Digital-to-Analog Converters

C. Output Devices

The output devices used with the analog computer must transcribe the voltage variations of the output into a form usable by the computer user. The method used depends upon whether the computer study being made is an on-line or off-line operation.

1. Control Systems

When the analog computer is used in an on-line application, as in a process control, the output voltage variation often is used to operate a control element in the system. For example, in a physiological support system where the computer output is proportional to the flow of oxygen into the subject's environment, the pressure at which the oxygen enters the system is controlled by the computer output by means of electrically operated valves and pumps driven by the computer.

2. Recording Applications

For off-line computations derived from on-line data inputs, the results of the computation may be retained by recording the computer output voltages on either paper graph recorders or magnetic tape recorders. These recording devices are discussed more fully in Sections III and IV.

3. Oscilloscope and Other Visual Readout Devices

If no permanent record of the computation is necessary, the computer output may be observed on an oscilloscope connected to the computer. This transitory method of viewing the computer output may suffice for initial alignment and calibration of the computer. After this point, a permanent recording of the actual computation may be desired.

D. Special Devices

1. Multipliers

In mechanizing various equations and functions on the analog computer, two variables often must be multiplied. The operational amplifiers of the analog computer are inadequate in this respect; various schemes have been devised to overcome this shortcoming.

Servo multipliers often are used to multiply. The position of a wiper arm on a potentiometer is proportional to a function of one variable, while the voltage across the ends of the potentiometer is proportional to the second variable. If either or both the wiper arm position or the voltage across the potentiometer vary, the output of the circuit across the wiper arm will vary, since it is proportional to the product of the two variables. Figure 134 illustrates a servo multiplier with provisions only for the multiplication of two positive numbers. Using additional power supplies, the device may be made to multiply both positive and negative quantities (so called four-quadrant multiplication). The position of the wiper arm is controlled by a servo-motor, whose angular position is proportional to one variable. The motor has some inertia and the wiper arm of the potentiometer has some friction, limiting the speed of the servo system to respond to a change in one quantity. Frequency variations of about 15 cycles per second are the maximum most of these units can follow with sufficient fidelity. The potentiometer also is limited in its accuracy because of loading by following stages. High input impedance operational amplifiers minimize this effect.

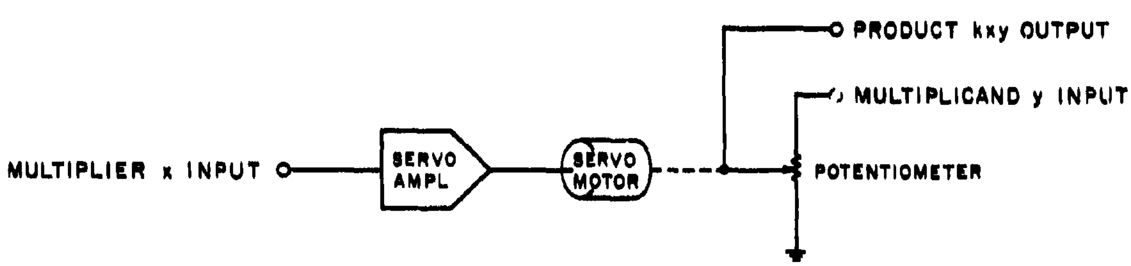


Figure 134. Servo Multiplier

A second type of multiplier commonly used is the time division electronic multiplier. In this device, the multiplication function depends on the fact that the average value of a pulse is proportional to both the amplitude of the pulse and the duty cycle of the pulse, i.e., the ratio of the duration of the pulse to the period of the pulse train. The multiplier controls the amplitude of the pulse train, the multiplicand controls the duration of the pulse, and the average value is proportional to the product. As long as the switching rate of the generator creating the pulses is much faster than the variation in the input signals, there may be as little as 0.01 percent error in this electronic multiplier. The device uses no mechanical action to limit the response to high frequencies, allowing inputs to vary as much as 2000 times per second.

A third type of multiplier is the quarter squares multiplier, which is based on the mathematical relationship:

$$P = xy = 1/4 \left[(x+y)^2 - (x-y)^2 \right] .$$

Since operational amplifiers can add and subtract quantities, the only problem in mechanizing this equation is the squaring function required. By using a biased diode function generator (see below), the squaring process is obtained easily and the quarter square multiplier is mechanized. It has practically an unlimited frequency response and an accuracy of better than 0.1 percent.

A fourth type, the log-antilog method of multiplying, is based on the fact that the sum of the logarithms of two numbers is the logarithm of the product of the numbers. In this device (shown in figure 135), the logarithms of the multiplier and multiplicand are generated using diode function generators (the voltage-current relationship of a diode is a very good approximation of a logarithmic response over a small range). The two logarithms are added with the aid of an operational amplifier, and then the antilog of the sum is derived, supplying the desired product.

There are other types of multipliers, but most of these have either accuracy or frequency limitations. The four described above are the most generally used.

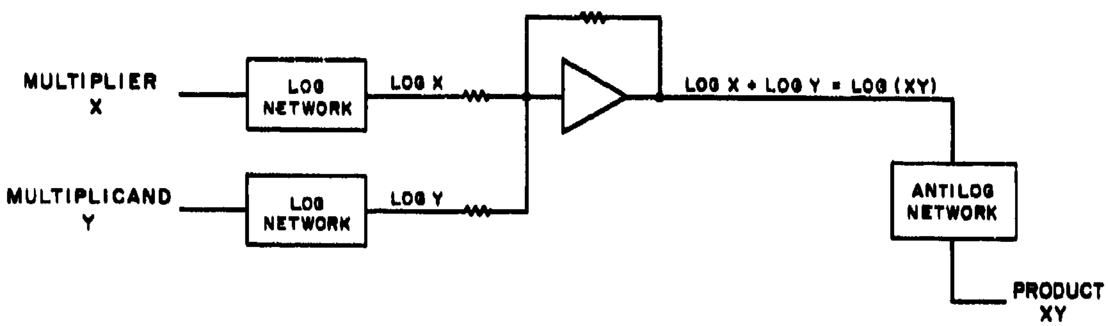


Figure 135. Log-Antilog Multiplier

2. Function Generators

A function generator is used to generate a voltage waveform that cannot be formed using only operational amplifiers and multipliers. Typical functions include the square of a variable, the square root, the logarithm, and the exponential. There are many types of function generators used with the analog computer. Depending upon the form in which the function is available, the method of generating an electrical signal equal to the function to be generated can range from simple manual plotting on a resistance paper to line approximations using biased diodes.

The usual method of simulating a nonlinear function is by approximating the voltage profile with straight line segments. Figure 136(A) illustrates an example of this approximation. A parabolic function ($y = x^2$) is generated with the use of diodes. The diodes are arranged in a circuit in such a way that both the slope and the breakpoint of the diode can be adjusted. Figure 137 shows a typical circuit for the adjustment of the breakpoint (the voltage at which the diode starts to conduct) and the slope (the change in output voltage per change in input) (ref. 46). Each diode has a conduction interval used for one of the straight lines in the approximation. By connecting the diode circuits together, as shown in figure 138, the straight line segments combine to form the parabolic function conductance. Actually, the gains of the operational amplifiers are changed as each diode conducts by changing the value of the input resistance. Figure 136(B) is the straight line approximation of the waveform, using segments formed with the diode function generator.

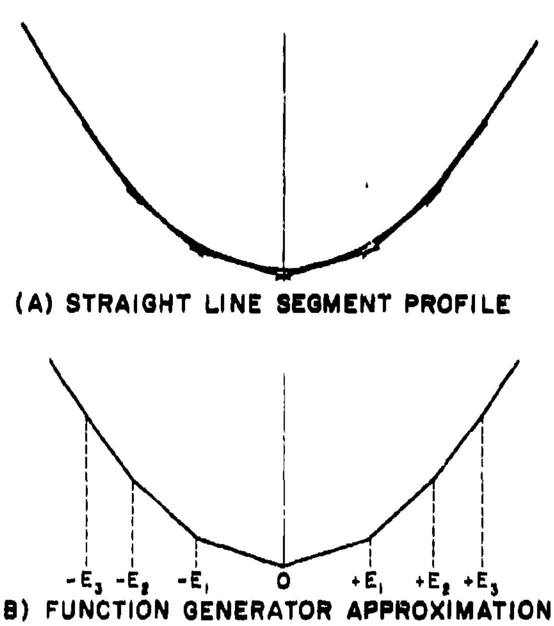


Figure 136. Diode Function Generator Operation

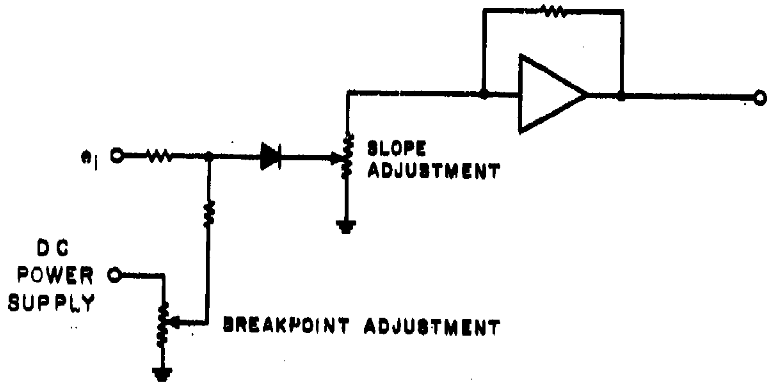


Figure 137. Diode Function Generator Circuit

If the function is plotted on paper or is in some nonelectrical form, it often is convenient to use the plotted waveshape to generate an electrical output that follows the variations in the function. (Devices discussed under digital computer analog-to-digital converters aid in this transformation.)

A tapped potentiometer may be used to approximate a functional relationship (see figure 139). By forcing the voltages at each tapped point to be a value equal to the power supply voltages, the function is approximated by straight line segments. Loading effects can deteriorate the accuracy of waveshape approximation.

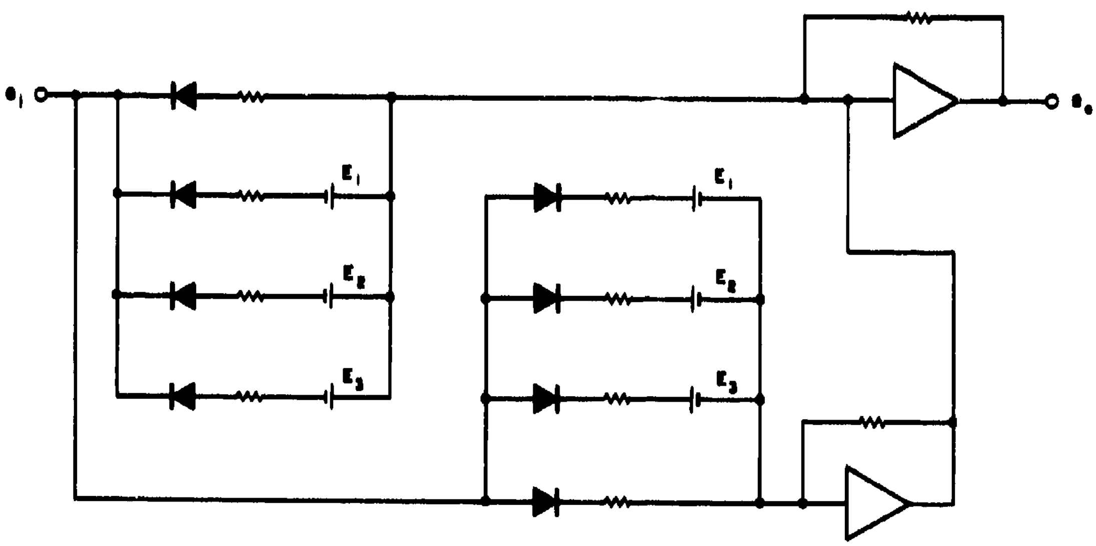
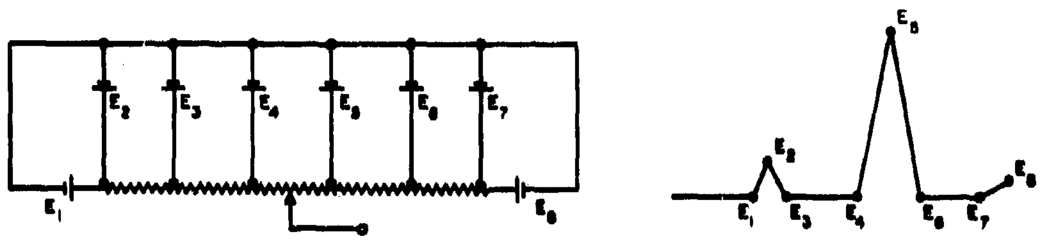


Figure 138. Parabolic Diode Function Generator



(A)
TAPPED POTENTIOMETER CIRCUIT

(B)
FUNCTION GENERATOR OUTPUT

Figure 139. Tapped Potentiometer Function Generator

Plotted values of the function also are used in conjunction with a photoelectric sensing system. The edge of an opaque pattern copied from the charted function is traced by a photoelectric follower system. The light on the curved edge may be supplied either from an ordinary lamp or from the light supplied by the face of a cathode-ray tube whose beam is being driven by a voltage proportional to the variable.

Various switching techniques coupled with biased diodes permit the introduction of parameters such as delay and hysteresis into an input function. Other nonlinear techniques include trigonometric function generation, including tapered potentiometers and a-c resolvers. These latter methods, however, require a servomotor to drive either the potentiometer or resolver.

3. Power Supplies

The internal power supplies used in the analog computer must be regulated very accurately and have essentially no superimposed noise (ripple or spikes). These power supplies are used to impose the initial conditions on the computing circuits when the computer is being used to derive solutions of differential equations and to simulate physical systems. The power output requirement for these supplies is not great, because their primary task is to charge a capacitor initially, after which the supplies are disconnected. Generally, there are potentiometers associated with the power supplies with switches so that the voltage needed may be obtained easily without additional patching between the power supply output and the potentiometer.

II. Operation

A. Interconnection of Components

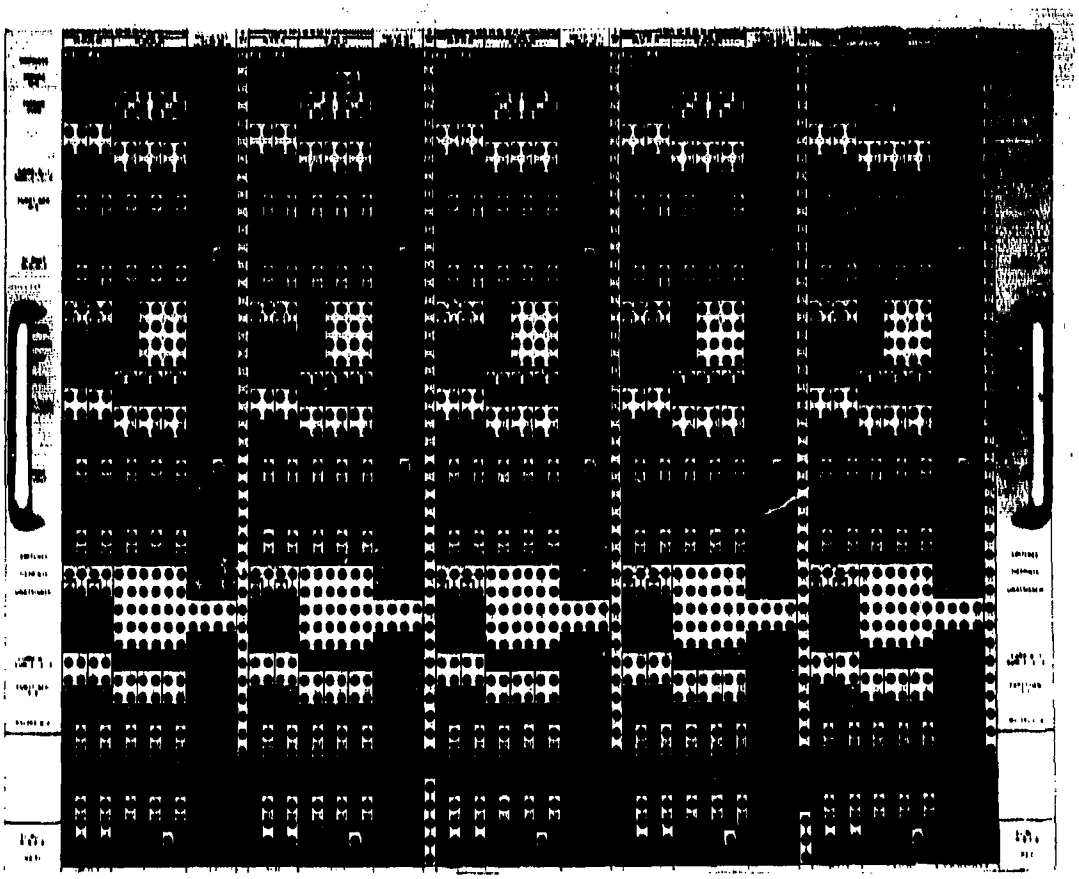
The components of the analog computer, the operational amplifiers, resistors, capacitors, potentiometers, power supplies, function generators, relays, and diodes, usually are connected together by a system of patchcords. With this system, the versatility and flexibility of the computer are not compromised by permanent connections.

The same operational amplifier can be connected as an integrator in one problem or as a multiplier in the following one. (There are special-purpose analog computers that are wired permanently to provide a certain function, such as a navigation computer.)

In general-purpose analog computers, connections to amplifiers, resistors, capacitors, potentiometers, and power supplies are available at a detachable patch-plug board. With this arrangement, a problem may be wired once, and then removed from the computer to allow another problem to be set up. When the first problem needs to be rerun, the patchboard already is set up and the computation can proceed with a minimum of preparation. A patchboard is shown in figure 140.

B. Normalizing

When components of an analog computer are connected, the quantities involved must be normalized so that no voltage variable exceeds the specifications of the individual stages. Normalization involves amplitude normalizing and time normalizing.



Electronic Associates, Inc., Long Branch, N.J.

Figure 140. Analog Computer Detachable Patchboard

1. Amplitude Normalizing

The input and output ranges of most operational amplifiers are ±100 volts, as noted above. If, however, a 50-volt output of one amplifier must be multiplied by a factor of 10 in the next stage, the following amplifier stage would be required to handle 500 volts, well beyond its capabilities. Therefore, the output of the first stage must be scaled down so that the second stage does not overload. Thus the 50-volt output might be scaled down by a factor of 10 to 5 volts, and then multiplied by 10 as required, which keeps the level within the 100-volt maximum. Of course to obtain the proper result in the output, the output voltage of the second stage must be considered to be 1/10 of the actual value of the computed variable.

In addition to scaling down a quantity to avoid overloading a stage, a quantity sometimes must be multiplied to improve the signal-to-noise figure in the calculation. Using the same example as above, if the output of one stage is at a 50-volt level and the next mathematical process to be performed is a division by 100, the signal level then would be reduced to 1/2 volt. While the computer should be linear and quiet at this level, it is preferable if the signal were more in the midrange of the 100-volt amplifier scale. Therefore, instead of dividing by 100, the computer would divide only by 10, or not divide at all, and the output is larger than the actual computed output by a factor of 10 or 100.

2. Time Normalizing

When a problem is not being computed on-line, the time of computation need not be the same as the real time associated with the problem. A computer time is defined (ref. 36) in some multiple or fraction of the time of the problem because of the following reasons:

a. Short runs usually have high frequencies, restricting the accuracy obtained with multipliers and demanding wide-band amplifiers.

b. The high frequencies associated with short runs may not be able to be reproduced adequately on the output recording instrument.

c. Short runs usually require higher amplifier gains.

d. Long runs introduce errors in integration.

e. Long runs require low value potentiometer settings, which are more liable to error.

The phase shift present in amplifiers usually increases with frequency, setting an upper limit of about 10 cycles per second as the maximum frequency used in the computation. Therefore, if the fundamental frequency of an oscillatory function

is over 10 cycles per second, the computation should be slowed down by a time factor. However, if the variation of parameters were so slow as to cause the computation time to run into hours, then the computation is scaled up by a time factor.

C. Initial Conditions

When computing the behavior of a dynamic system (a typical problem for the analog computer), the conditions of the system when the computation begins must be stated. These initial conditions, which may be the initial position of some mass or its initial velocity, or the initial flow or pressure in a hemodynamic system, are represented by voltages impressed across the amplifier inputs or outputs before the computation begins, forcing the function to be a certain value at time zero. The internal power supplies of the computer are needed for this application.

Usually, the initial condition voltage is impressed across the capacitor in the integrating circuit, which makes it imperative that the capacitors be of high quality, have low leakage, and be stable with time. Often the capacitors, as well as the resistors, are kept in a thermostatically controlled chamber to minimize changes because of temperature variations. After the computation starts, the initial conditions are removed from the computer by relay controls.

D. Sample Problems

A typical problem for the analog computer is the solution of a differential equation representing the interactions in a physiological control system. This mathematical description of the physics of a system is common to all fields of science, from electronics to physiology. A typical problem is a second-degree differential equation (linear with constant coefficients) of the form

$$a_2 \frac{d^2y}{dt^2} + a_1 \frac{dy}{dt} + a_o y = Au_o(t)$$

which describes a damped sinusoidal oscillation generated by a step change input to the system. Rewriting the equation to solve for the highest order derivative:

$$\frac{d^2y}{dt^2} = A/a_2\, u_o\,(t) - a_o/a_2\, y - a_1/a_2 \frac{dy}{dt}$$

The mechanization for this problem is shown in figure 141. If the quantities A/a_2, a_o/a_2, and a_1/a_2 cause the three quantities of which they are coefficients to be greater than 100, then amplitude scaling is necessary. It can be shown that $\sqrt{a_o}$ is the natural frequency of the system, and if this quantity is over 10 cycles per second, then time scaling may be necessary.

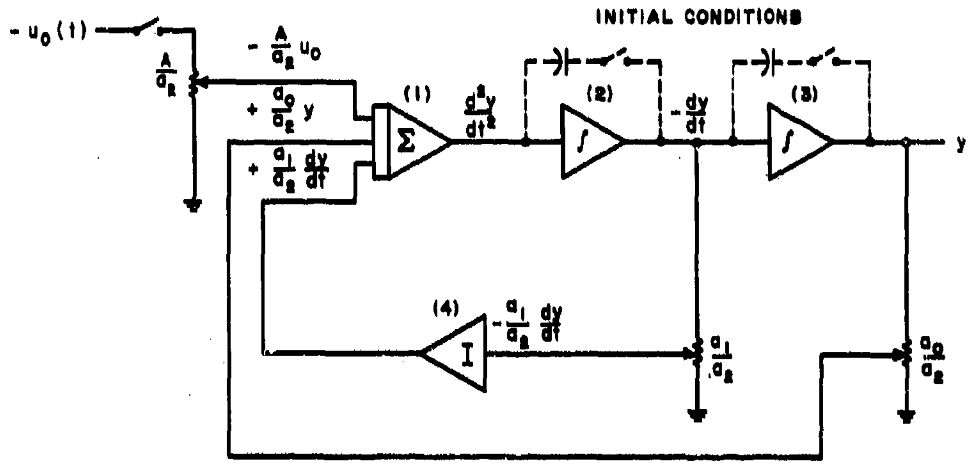


Figure 141. Solution of a Sample Problem on an Analog Computer

The first operational amplifier (1) is simply an adder circuit, adding the three input quantities to form d^2y/dt^2 as defined in the above equation. The next amplifier (2) is connected as an integrator (as well as inverting), forming $-dy/dt$; the third amplifier forms y by integrating again. To form the inputs for the first amplifier, the coefficient potentiometers are set, and an inversion is performed in the fourth amplifier to change $-dy/dt$ to $+dy$ dt, which is required. Since the amplifier (1) performs an inversion, the algebraic sum is correct. As the problem is mechanized, the initial conditions are considered to be zero; i.e., at time $t = 0$, $y = o$ and $dy/dt = 0$. If this were not the case, then voltage sources of the right magnitude and polarity are connected across the appropriate amplifiers to force the proper initial conditions. At the time the computation starts, these voltage sources are disconnected by the relay contacts shown.

DATA PROCESSING APPLICATIONS IN PHYSIOLOGICAL MONITORING

The use of automatic data processing in the field of physiological monitoring has been somewhat limited. This has been due, in part, to the unavailability of suitable equipment, and to the physiologist's ignorance of suitable techniques. More important, perhaps, has been the simple lack of knowledge: in many instances, normal responses of the human to environmental stress are unknown. Further, all the psychological, physiological, and environmental factors bearing on a measurement situation must be known and must be categorized for effective machine operations.

Historically, data analysis has been accomplished by a clinician or physiologist, using what precise measurements he could obtain, and comparing them to past records, measurements, and case histories. His determinations are often subjective, being made on knowledge of contributing factors which is largely unquantified. The major task in attempting to automate the processing of physiological data is the accumulation and

categorization of sufficient quantitative data so that the fast but relatively simple operations of data processing equipment can accomplish the sophisticated but slow comparison of the trained clinician.

Most activity in this field to date has been in two areas where a relatively large store of such data exists: electrocardiography and electroencephalography.

1. Electrocardiography

Data processing techniques for the analysis of the electrocardiogram generally are divided into two phases: the preparation of the data for the automatic analysis and the analysis itself.

A. Data Preparation

Obviously, the process for preparing data differs depending on the type of computer that is used. If the automatic data processing technique involves the use of an analog computer, the data must be in a form compatible with the input requirements of the computer. For an ECG, this involves either recording the original ECG and slowing down the playback recording, or else filtering the ECG to remove the higher frequencies, because the ECG has a broader frequency band than most analog computers can handle adequately. The recording-and-playback-at-slower-speed method often is superior because no information is lost, as opposed to the filtering technique. The final decision depends upon what parameter of the ECG is of real interest.

If the data are being prepared for digital computer analysis, the first step is to digitize the incoming ECG. With a bandwidth of 200 cycles per second considered significant, a sampling rate of 1000 samples per second was used by one group (ref. 53). This digitized form of the ECG waveform then was converted by an analog-to-digital converter to a 36-character word, seven characters per lead representing the numerical value of the ECG and the remaining characters used for identification purposes. This form of the ECG was available for computer analysis, using the IBM 704 Data Processing System. This format is typical of most arrangements for digital computer processing of the information.

B. Data Analysis

The method of analyzing an ECG with automated techniques is a decision to be made by the experimenter based on the use he intends to make of the information. The duration of the P, QRS, and T segments of the ECG waveform is one parameter of interest for some researchers. Either analog (ref. 20) or digital (ref. 39) techniques may be used. Other researchers have investigated the correspondence between ventricular gradient characteristics and cardiac irregularities (ref. 38). Both analog (ref. 29) and digital (ref. 53) techniques have been used to determine the polar vector for use in diagnosis and standard categorization of ECG signals. Other work has included the extraction of standard ECG waveshape parameters (ref. 8).

II. Electroencephalography

A. Data Preparation

In general the preparation of EEG data is similar to that for ECG. A sampling rate of 45 kc to form binary coded information was used by one group (ref. 56). Smoothing the data by passing the waveshape through a low-pass filter before digitizing can eliminate some of the noise superimposed on the EEG and reduce the necessary sampling rate.

B. Data Analysis

As in the case of the ECG, the analysis of EEG information using automated techniques depends greatly upon the researcher's goals. Programs for EEG pattern recognition have been constructed by numerous workers in the field (ref. 7, 13, 43). Digital computer techniques also are used to separate the effects of environment on the observed EEG from the intrinsic EEG patterns caused by stress, fatigue, and other mental reactions (ref. 18). An involved time series analysis performs a mathematical mapping of the EEG responses in such a way that it is possible to distinguish the effects of fatigue from among the normal variations and noise present in the EEG.

Although most of the methods of analysis involve the use of the analog or digital computer, some simpler means of data reduction and analysis are available. One method (ref. 28) simply involves the punching on cards of the collected EEG data for ease in future diagnosis and clinical testing.

III. Problems in Processing of Physiological Data

There are a few problems associated with data processing techniques which, if not recognized, can either decrease the worth of the data obtained, or in the extreme, render the data useless while consuming time and money. In most cases these problem areas lie in the realm of initial preparation of the data for computer use after having obtained the raw data.

A. Sampling and Digitizing

When the input data are being prepared for digital computer computation, errors can occur when the data are digitized. The rate at which the data are sampled for digitizing is a function of the highest frequency components contained in the data. If f^* is the highest frequency contained in a composite signal, the sampling rate must be greater than $2f^*$ for unambiguous results (ref. 22). Generally it is preferred to keep the sampling rate as low as possible. The faster the sampling rate, the more memory locations will be necessary to store the digitized data. If there are high-frequency components that are not of interest, they must be filtered out before the sampling takes place, if a lower sampling frequency is to be used. Sampling cannot take the place of filtering.

The analog-to-digital converter has only a finite accuracy. A knowledge of this accuracy is essential if the results obtained from the data computation are to be trusted and used. The accuracy is determined by the value of the least significant bit.

B. Editing

If the results of an experiment produce great quantities of data, there is the problem concerning the disposition of the data. Either all data are entered into the computing system and the computer and programmer extract the significant portions, or the data must be pre-edited before being entered into the computational device. If all data are entered, then sufficient storage space within the computer must be assured, and sufficient time must be allowed to sort and edit the information. Time on a digital computer is relatively expensive; however, man is relatively slow when called upon to sort the immense quantities of information involved. Therefore, the most efficient and economical manner of data retrieval must be chosen carefully.

C. Calibration

Because of variations in both the electronic equipment associated with data recording and the values of physiological parameters associated with biological signals, some means of calibration must be included with the data input procedure to insure the standardization of computations. This procedure may be as simple as including a calibration signal of some known amplitude at the beginning and end of the input information. The automatic data processor adjusts the digitizer or converter to normalize the voltage for the use of the computer, as well as indicating the normalization factor so that the true amplitude can be known.

D. Standardization of Data Form

There are many different computer systems used in different facilities. Each system generally has its own unique format in digitizing information for computational purposes. Within a given facility, this causes no problem, but when information and data are to be exchanged among facilities, or if a centralized computer facility processes input data from many collection agencies, the need for standardized formats is imperative.

In digital tape formats, the standardization specifications include recording density, parity check convention, and the length of the blank tape between adjacent sets of data. Other digital conventions include the standardization of punched card codes, sizes of punched cards, punched paper tape codes, widths of punched paper tape, and formats of information on the paper tape.

Within the recording of the parameters themselves standardization is necessary in lead placement, electrode placement, and other physical factors which influence the interchangeability of data between experimenters.

BIBLIOGRAPHY

1. Aerospace Instrumentation Handbook, Air Force Flight Test Center, Edwards AFB, Calif., 1962.

2. Armstrong, Harry G. (Editor), Aerospace Medicine, Williams and Williams Co., Baltimore, Md., 1961.

3. Basic Electronics, Navpers 10087, U.S. Government Printing Office, Washington 25, D. C., 1955.

4. Begun, S. J., Magnetic Recording, Rinehart and Co., Inc., New York, N. Y., 1958.

5. Borden, Perry A., and Wilfred J. Mayo-Wells, Telemetering Systems, Reinhold Publishing Corp., New York, N. Y., 1959.

6. Brigden, J. K., "Input-Output Equipment for Digital Computers," in Handbook of Automation, Computation, and Control, Vol. II, John Wiley and Sons, Inc., New York, N. Y., 1959.

7. Burch, N., A. Silverman, and T. H. Greiner, "New Approach to Theory and Automatic Analysis of EEG," Electroencephalogram and Clinical Neurophysiology, Vol. 8, p. 157, 1956.

8. Caceres, C. A., et al, "Computer Extraction of Electrocardiographic Parameters," Circulation, Vol. 25, p. 2, 1962.

9. Chance, Britton, et al (Editors), Electronic Time Measurements, McGraw-Hill Book Co., Inc., New York, N. Y., 1949.

10. Cherry, Colin, On Human Communication, John Wiley and Sons, Inc., New York, N. Y., 1957.

11. Donaldson, P. E. K., et al, Electronic Apparatus for Biological Research, Butterworths Scientific Publications, London, England, 1958.

12. Elsner, R. W., C. J. Eagan, and S. Anderson, "Impedance Matching Circuit for the Mercury Strain Gage," Journal of Applied Physiology, Vol. 14, p. 871, 1959.

13. Farley, B. G., et al, "Computer Techniques for the Study of Patterns in the Electroencephalogram," IRE Transactions on Bio-Medical Electronics, Vol. 9, p. 4, 1962.

14. Geddes, L. A., E. H. Hoff, and W. A. Spencer, "Short Distance Broadcasting of Physiological Data," IRE Transactions on Bio-Medical Electronics, Vol. 8, p. 168, 1961.

15. Glasser, Otto (Editor), Medical Physics, Year Book Publishers, Inc., Chicago, Ill., 1944 (Vol. I), 1950 (Vol. II), 1960 (Vol. III).

16. Goldberg, M. N., R. A. Mills, and W. V. Blockley, Instrumentation Package for In-Flight Physiological Studies, WADD TR 60-83, Wright Air Development Division, Wright-Patterson AFB, Ohio, 1960.

17. Hershberger, W. D., Principles of Communications Systems, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1955.

18. Himmelman, D. S., "Electroencephalographic Time Series Analysis," presented at 5th International Conference on Medical Electronics, Liege, Belgium, July 1963.

19. Holzbock, Werner G., Instrumentation for Measurement and Control, Reinhold Publishing Corp., New York, N. Y., 1962.

20. Johnson, F. D., R. McFee, and J. M. Bryant, "An Integrating Circuit for Measurement of the Areas of the Waves of the Electrocardiogram," Circulation, Vol. 2, p. 5, 1950.

21. Kret, David B., Transducers, Allen B. DuMont Laboratories, Inc., Clifton, N. J., 1953.

22. Ledley, R. S., Digital Computer and Control Engineering, McGraw-Hill Book Co., Inc., New York, N. Y., 1960.

23. Lion, Kurt S., "Mechano-Electric Transducer," Review of Scientific Instruments, Vol. 27, p. 222, 1956.

24. Lion, Kurt S., Instrumentation in Scientific Research, McGraw-Hill Book Co., Inc., New York, N. Y., 1959.

25. Lucchina, G. G., and C. G. Phipps, "A Vectorcardiographic Lead System and Physiologic Electrode Configuration for Dynamic Readout," Aerospace Medicine, Vol. 33, p. 722, 1962.

26. Marko, Adolf R., Monitoring Unit for Heart and Respiration Rate, WADD TR 60-619, Wright Air Development Division, Wright-Patterson AFB, Ohio, 1960.

27. Marko, Adolf R., M. A. McLennan, and E. G. Correll, A Multichannel Personal Telemetry System Using Pulse Position Modulation, ASD TR 61-290, Aeronautical Systems Division, Wright-Patterson AFB, Ohio, 1961.

28. Metcalf, D. R., H. D. Lockett, and E. S. Turrell, "A Flexible IBM System for Collating EEG and Clinical Data," Electroencephalogram and Clinical Neurophysiology, Vol. 12, p. 202, 1960.

29. Moore, A. D., P. Harding, and G. E. Dawer, "The Polarcardiograph," American Heart Journal, Vol. 64, p. 382, 1962.

30. Nelson, Robert C., "Recorder Survey," Instruments and Control Systems, Vol. 35, p. 75, July 1962.

31. Neville, J. Ryan, "Electrochemical Device for Measuring Oxygen," Review of Scientific Instruments, Vol. 33, p. 51, 1962.

32. Nichols, Myron H., and Lawrence L. Rauch, Radio Telemetry, John Wiley and Sons, Inc., New York, N. Y., 1956.

33. Nyboer, Jan, "Plethysmograph: Impedance," in Medical Physics, Vol. II, Year Book Publishers, Inc., Chicago, Ill., 1950.

34. O'Kelley, L. F., et al, Telemetry Transducer Handbook, WADD TR 61-67, Wright Air Development Division, Wright-Patterson AFB, Ohio, 1961.

35. Partridge, G. R., Principles of Electronic Instruments, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1958.

36. Pfeffer, I., "Linear Electronic Computer Elements," in Handbook of Automation, Computation, and Control, Vol. II, John Wiley and Sons, Inc., New York, N. Y., 1959.

37. Pierce, J. R., Symbols, Signals, and Noise, Harper and Bros., New York, N. Y., 1961.

38. Pipberger, H. V., "Data Processing for Electrocardiography," in Data Acquisition and Processing in Biology and Medicine, Pergamon Press, Inc., New York, N. Y., 1962.

39. Pipberger, H. V., R. J. Arms, and F. W. Stallman, "Automatic Screening of Normal and Abnormal Electrocardiograms by Means of a Digital Electronic Computer," Proceedings of the Society for Experimental Biology and Medicine, Vol. 106, p. 130, 1961.

40. Radar Electronic Fundamentals, Navships 900, 016, U. S. Government Printing Office, Washington 25, D. C., 1944.

41. Radio Amateur's Handbook, American Radio Relay League, West Hartford, Conn., 1963.

42. Results of the First U. S. Manned Orbital Space Flight, National Aeronautics and Space Administration, U. S. Government Printing Office, Washington 25, D. C., 1962.

43. Riehl, J. L., "Analog Analysis of EEG Activity," Aerospace Medicine, Vol. 32, p. 1101, 1961.

44. Roman, James A., and Lawrence E. Lamb, "Electrocardiography in Flight," Aerospace Medicine, Vol. 33, p. 527, 1962.

45. Schwartz, M., Information Transmission, Modulation, and Noise, McGraw-Hill Book Co., Inc., New York, N. Y., 1959.

46. Scott, N. R., Analog and Digital Computer Technology, McGraw-Hill Book Co., Inc., New York, N. Y., 1960.

47. Smedal, Harald A., G. R. Holden, and J. R. Smith, Jr., Flight Evaluation of an Airborne Physiological Instrumentation System, Including Preliminary Results Under Conditions of Varying Accelerations, NASA TN D-351, Ames Research Center, Moffett Field, Calif., 1960.

48. Smith, Stewart L., "Magnetic Tape Recording," in Aerospace Telemetry, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1961.

49. Sosnow, M., and E. Ross, Electrodes for Recording Primary Bioelectric Signals, ASD TR 61-437, Aeronautical Systems Division, Wright-Patterson AFB, Ohio, 1961.

50. Stacy, Ralph W., Biological and Medical Electronics, McGraw-Hill Book Co., Inc., New York, N. Y., 1960.

51. Stiltz, Harry (Editor), Aerospace Telemetry, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1961.

52. Suckling, E. E., Bioelectricity, McGraw-Hill Book Co., Inc., New York, N. Y., 1961.

53. Taback, L., et al, "Digital Recording of Electrocardiographic Data for Analysis by a Digital Computer," IRE Transactions on Medical Electronics, Vol. 6, p. 167, 1959.

54. Telemetry Standards, IRIG Document No. 106-60, Inter-Range Instrumentation Group, White Sands Missile Range, New Mexico, 1960.

55. Tolles, Walter E., and William J. Carberry, A System for Monitoring the Electrocardiogram During Body Movement, WADC TR 58-453, Wright Air Development Center, Wright-Patterson AFB, Ohio, 1959.

56. Tunturi, A. R., and L. W. Bagby, "Statistical Data Acquisition and Processing System for Medical Research," presented at 4th International Conference on Medical Electronics, New York, N. Y., 1961.

57. Veghte, J. H., Human Physiological Response to Extremity and Body Cooling, AAL TR 61-26, Arctic Aeromedical Laboratory, Ft. Wainwright, Alaska, 1961.

58. Weber, Paul J., The Tape Recorder as an Instrumentation Device, Ampex Corp., Redwood City, Calif., 1959.

59. Wheelwright, Charles D., Physiological Sensors for Use in Project Mercury, NASA TN D-1082, Manned Spacecraft Center, Houston, Texas, 1962.

60. Wood, E. H., W. F. Sutterer, and L. Cronin, "Oximetry," in Medical Physics, Vol. III, Year Book Publishers, Inc., Chicago, Ill., 1960.

INDEX